Individual differences in foreign language attainment of children with poor literacy skills

Alexa von Hagen

A thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

ARC Centre of Excellence in Cognition and its Disorders (CCD)

Department of Cognitive Science

Macquarie University, Sydney, Australia

August 2018

The work presented in this thesis was carried out as part of the Erasmus Mundus joint International Doctorate for Experimental Approaches to Language and Brain (IDEALAB) and was funded by the European Commission within the action nr. 2015-1603/001-001-EMJD (Framework Partnership Agreement 2012-2025).

Macquarie University (Australia)

Newcastle University (United Kingdom)

Rijksuniversiteit Groningen (the Netherlands)

Università degli studi di Trento (Italy)

Universität Potsdam (Germany)

Table of Contents

Summary

A common belief in many educational contexts is that students with poor literacy skills are worse at learning foreign languages than their peers with typical literacy skills. Past research confirms this concern. However, this evidence is almost entirely based on group studies, which average highly variable performance of individual children. A systematic investigation of individual differences in the foreign language attainment of children with poor literacy skills is currently missing. The present project aimed to fill this gap.

The first chapter is a systematic review and meta-analysis of the available evidence on this topic. The second chapter investigated to what extent results of group comparisons between children with poor and typical literacy skills are representative of the individual foreign language performance of poor readers/spellers. The third chapter explores potential sources of this variability, with a focus on native language skills. Finally, in the fourth chapter, foreign language word knowledge is examined as a function of task complexity and word characteristics.

In summary, this project deepens our understanding of individual differences in foreign language attainment of children with poor literacy skills. Our findings reveal that the common belief that children with poor literacy skills are worse at learning foreign languages than their classmates with average literacy skills, does not hold for every child with poor literacy skills. Even when foreign language learning difficulties emerge, the subskills in which poor readers/spellers struggle are quite diverse. The extent to which these foreign language deficits are associated with equivalent native language deficits seems to vary across subskills. Overall, the results of this project provide new insights to guide parents and teachers in making evidence-based decisions on the foreign language education of children with poor literacy skills.

Statement of Originality

The work in this thesis is my original work. It has not been submitted for a higher degree in any other university or institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University and the University of Potsdam in the International Doctorate of Experimental Approaches to Language and Brain (IDEALAB). Ethics approval for the research reported in this thesis was obtained from the Human Research Ethics Committee (Human Sciences and Humanities) at Macquarie University (reference no 5201600544) and the Ethics Committee at Potsdam University (reference no 49/2015).

Alexa von Hagen

Student number: 44795890

31st of August, 2018

Acknowledgements

"The limits of my language are the limits of my world."

Ludwig Wittgenstein

A thousand times thank you, Danke and merci to my supervisors Saskia Kohnen and Nicole Stadie for being the most amazing supervisors I could ever have whished for. Three years of endless patience, constant support, encouraging words, stimulating conversations, challenging questions, gentle advice, dynamic teamwork, many many Skype meetings and so much learning. You are my absolute superheroes!

Thank you, Grazie, Dank U and vielen Dank to Prof. Lyndsey Nickels, Prof. David Howard, Prof. Gabriele Miceli, Prof. Roelien Bastiaanse and Prof. Barbara Höhle for giving me the opportunity to join such a wonderful program like IDEALAB. Thank you for engaging with our research projects during every winter- and summerschool and also for taking time to enjoy many social events with us.

Thank you so much and vielen vielen Dank to the children, parents and teachers that made this work possible. Especially Pamela Morio and Elizabeth Blair gave me such a kind and warm welcome to their classrooms and schools. Vielen lieben Dank also to Elena Högl and Hannah Pranaitis for your careful dedication during data collection! This thesis would not have been possible without you.

A very special Dankeschön to Caro Schröter and Jenny von Frankenberg from the Zapp Berlin, who were generously taught me how poor literacy skills are identified and treated in Germany. I learned so much from you! Vielen lieben Dank also to Sandra Hanne, Tom Fritzsche and Anne Adelt from Potsdam University for your support during the planning and implementation of my research. Tausend Dank also to Anja Papke for tolerating all of my questions and always answering with a kind smile.

At Macquarie University, I especially want to say thank you and merci to Serje Robidux, who very patiently answered all my questions about statistical analyses and made

me feel confident that I was doing the right thing. Also thank you so much to Deanna Francis, Nick Badcock and John Elias for sharing your knowledge on systematic reviews with me. To Lesley McKnight, thank you for your cheerful and warm welcome and for taking so good care of us Idealabers during the winterschool.

Thank you, Whakawhetai ki a koe, Grazie milliano, Tausend Dank, cảm ơn nhiều lắm, Maraming salamat, Dank je wel, baie dankie to Ella, Ingaleinchen, Hanh, Wengelchen, Jakolien and Steph! You have been such an important part of my life in these last three years and I can't wait to see where we will re-unite next and celebrate all together! A very special muito obrigada to Aninha! I´m already missing your wise advice with our Machiatti at Café al Teatro. Also Grazie mille to Prerana, Leonie and Leanne for all those enjoyable Sydney and Rovereto lunch breaks! Grazie mille also to Fleur, for passing on all your tips on fleuring around. Very important contribution to these three years! I also want to thank Hui-Ching, Assunta, Nenad and Serine for making sure I also extended my culinary knowledge during my time in IDEALAB. Very much appreciated!

There are also many people that had a direct influence on my decision to pursue a PhD. First, from my Masters program at Edinburgh University, thank you, Ευχαριστώ πολύ, Grazie mille and ありがとう to Vicky Chondrogianni, Antonella Sorace and Mits Ota. You would not imagine how many times I returned to the materials of your lectures during these three years. Second, millones de gracias to Sylvia Defior and Paqui Serrano from the University of Granada. This whole academic journey abroad started when you kindly received me in beautiful Andalucía. Last, but certainly not least, I want to say infinitas gracias to my colleagues and friends from the Universidad Católica del Uruguay. ¡Qué lujo poder trabajar con profesionales tan brillantes como Cindy Mels, el Dr. Nacho Navarrette, Daniel Costa, Andrea Roca, Rosario Aldama, Leticia Vázquez, Daniel Trías, Cecilia Cracco y tantos más! Sigo teniendo muy presente todo lo que aprendí con Ustedes. Y finalmente, sin palabras el agradecimiento a mi gran profesora Alejandra Balbi y al Dr. Ariel Cuadro. Sin duda son los

principales responsables de que haya emprendido este camino hacia el mundo académico. Muchísimas gracias por siempre apoyarme y hacerme sentir, que a pesar de la distancia, sigo siendo parte del mundo UCU.

¡Mil millones de gracias también a mis amigas psicopedagogas! Lula, Deni y Mecha: Gracias por ser mi cable a tierra a la realidad clínica con nuestros eternos audios por whatsapp y por seguir con esa misma sed por aprender, que nos unió hace tantos años en CEDIIAP. Carola, Sofi, Clara, Nati y Analía: Gracias por dejarme seguir siendo parte, de todas esas charlas sobre la formación de las futuras generaciones de nuestra carrera. A Jime: Gracias por esa amistad tan especial, que nos unió hace tantos años en torno a la psicopedagogía.

Thank you also to my sister Merve, who was a very important support during these three years. Thanks to you and your family I had quite a special 'student' experience here in Sydney! Re-connecting with you was one of the highlights of these three years. También quiero agradecerle muy especialmente a mi querida Titi, que fue quien me contó hace muchos años, que existía una carrera llamada psicopedagogía. ¡Mil gracias Titi por horas y horas de charlas sobre lo que tanto nos gusta! Un riesengrosses Dankeschön y mil gracias a mi hermano Fede, que siempre me viene a visitar, esté donde esté en el mundo. Gracias por ser tanto hermano mayor como hermano menor, con una mezcla justa de seriedad, consejos, risas y complicidad.

Finally, but most importantly, I want to thank my parents, who from childhood onwards taught me the two most important lessons of my life: to feel at home anywhere in the world and to be ready to turn the page and be excited about the new chapter that is starting in my life. ¡Muchas muchas gracias! Realmente esos dos aprendizajes fueron la clave para disfrutar al máximo de IDEALAB.

General introduction

Imagine if you asked a random group of people to share their general thoughts on providing foreign language instruction in primary school. What would they say? Somebody might say that children should learn a foreign language as early as possible, because they will need to be able to speak another language fluently if they want to work in a well-payed job in the future. Someone else might highlight how important it is for children to be able to communicate in several languages, a skill that is becoming more and more important in today's increasingly multilingual societies. Another person might say how much easier it would be for their children to engage with locals on their next vacation and get more authentic experiences of the host country's culture. Maybe somebody also mentions that learning a foreign language is important, because you gain knowledge about language in general and as a consequence, also learn more about your own native language.

What would happen, if you then asked the same people a second question: What about a child who is already struggling at school, for example they are having difficulties in learning to read and write? Should they be exposed to foreign language classes? Some people would probably express concerns and say the advantages of foreign language learning may not be as important for children who have been identified with poor literacy skills in their native language. Somebody might say that children with poor literacy will probably not be as successful as their classmates in learning a foreign language. Someone else might see foreign language instruction as an additional burden for children with poor literacy skills and would recommend that they first strengthen their native language, before adding another challenge. Another person might also believe that learning a foreign language can have detrimental effects for children with poor literacy skills, because they might get "mixed up" between the two languages and start making more mistakes in their native language.

Opinions like these are very common among parents, teachers and specialists working with children with poor literacy skills (Baker, 1996; Genesee, 2015; Hussien, 2014; Palladino, Bellagamba, Ferrari & Cornoldi, 2013; Miller-Guron & Lundberg, 2000; Peer & Reid, 2014; Sparks, 2016; Wight, 2015). In this context, the British Dyslexia Association highlights that "dyslexic pupils learning a foreign language have to be aware that it may be a longer process for them than for others", although they emphasize that each student should still have the opportunity to receive foreign language instruction (British Dyslexia Association, 2016).

This common belief that poor readers will have trouble learning a new language has important implications, as it often leads to partial or even complete exemptions from foreign language instruction for children with poor literacy skills (Palladino et al., 2013; Sparks, 2016; Wight, 2015). This means that children with poor literacy skills are often not given the same opportunities as their classmates with typical literacy skills to gain many of the advantages associated with foreign language learning (e.g. advantage on inhibitory control - Bialystok & Majumder, 1998; advantage in theory of mind development – Kovács, 2009; advantage on metalinguistic knowledge – Bialystok, 2012). This would probably be justified if detrimental effects on mental health and native language skills were the consequence of exposing poor reader/spellers to foreign language instruction. Surprisingly however, decisions of exempting students with poor literacy skills from foreign language classes are often "(1) based on personal beliefs and preferences rather than on the basis of a carefully considered consensus of inclusion, and (2) in the absence of actual data about the potential successes of students with special needs" (pp.41-42, Wight, 2015). An evidence base to support decisions on the foreign language education of children with poor literacy skills is therefore urgently needed. The overall aim of this thesis was to contribute to this evidence base, by investigating individual differences in foreign language attainment of children with poor literacy skills.

To set the stage for the work presented in this thesis, in the following sections we define some of the central terms used in this work. First, we use the term 'poor readers/spellers' or 'children with poor literacy skills' to refer to students that have been identified with poor reading and/or spelling skills in their native language. More specifically, these difficulties encompass inaccurate or slow reading of words, nonwords, sentences or text and inaccurate spelling of words and nonwords (Elliott & Grigorenko, 2014; IDA, 2012; Siegel, 2018). Past research has used many different terms to refer to these difficulties, such as for example developmental dyslexia and/or dysgraphia, specific reading and/or spelling difficulty, reading and/or spelling impairment, deficit or disability (Elliott & Grigorenko, 2014; Siegel, 1988a, 2007; Serry & Hammond, 2015)[1].

Descriptors such as 'reading impairment' commonly serve as umbrella terms to refer to children with very heterogeneous cognitive profiles (Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013; Moll & Landerl, 2009; Zoccolotti & Friedmann, 2010). While all poor readers/spellers, by definition, experience at least one of the above-mentioned reading or spelling deficits, the specific reading and spelling skills affected in each child are different. Moreover, some, but not all children, also struggle with oral language and/or present other learning disorders (e.g. developmental dyscalculia, attentional deficit disorder, etc.) (Germanò, Gagliano & Curatolo, 2010; Landerl & Moll, 2010; McArthur, Hogben, Edwards, Heath & Mengler, 2000; Moll, Göbel & Snowling, 2015).

This heterogeneity is important. Depending on what exact difficulties a child shows, they may or may not show a lower foreign language attainment than their peers with typical literacy skills. Even when children do show difficulties, the heterogeneity implies that it is also likely that the foreign language subskills poor readers/spellers show difficulties which differ from child to child. Heterogeneity can be difficult to capture in group studies. Yet, available evidence on foreign language attainment in children with poor literacy skills is predominantly based on group averages that aggregate the scores of individual poor readers/spellers that are likely to have very different native language profiles (e.g. Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006). This means that we do not know if this evidence accurately

[1]In the introduction of Chapter 1 we provide a more detailed definition of the term 'children with poor literacy skills'.

reflects the foreign language profiles of individual poor readers/spellers and can therefore be used to guide decisions on the foreign language education of individual students. One important contribution of the work presented in this thesis is therefore to provide information on foreign language attainment of children with poor literacy skills that is representative of individual poor readers/spellers. We address these aspects in Chapters 1 and 2.

Moreover, we want to acknowledge that comparing children with poor and typical literacy skills, as we do in several of the studies reported in this thesis, is a rather crude comparison (MacCallum, Zhang, Preacher & Rucker, 2002). However, in practice, this dichotomy (poor vs. typical) is a reality. For instance, school psychologists or speech-language therapists often conduct literacy screenings or diagnostic assessments and are guided by arbitrarily set cut-off score. At the end of an assessment, children are either described as having a difficulty (below cut-off) or not (above cut-off). These dichotomous categories are also used in school regulations impacting on support strategies, assessment accommodations or exemptions (e.g., Schulte-Körne, 2010). Lastly, past research has predominantly focused on comparing children with poor and typical literacy skills on different foreign language measures (Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006). Although we are aware of the fact that literacy skills are continuously distributed, we believe that allocating children to a poor or typical literacy group allows us to address practical implications in an ecologically valid way. It also aids comparisons to previous research.

Turning now to the term ‘foreign language attainment’, we refer to the proficiency achieved in a language other than a person's native language through classroom based foreign language instruction[2]. More specifically, this means that in this thesis, we studied children who were exposed to the foreign language at school and had no additional exposure to the foreign language, other than limited access for example to music, films and computer games.

[2]Further details on different types of foreign language instruction are presented in the introduction of Chapter 1.

This is different to other situations in which children learn an additional language, for example through the exposure to more than one language from birth. Also, additional language learning through immersion in an environment in which this language is dominant (e.g. when a family moves to a new country, children of immigrant families or so-called 'English-language learners') is not the subject of this thesis. Although we draw on several of the theoretical contributions provided by past research on different types of child bilingual populations, the evidence obtained by studies in this field are not suitable to address the questions that we ask in this thesis. The reason for this is that foreign language learners and other bilingual groups have important differences in the quantity and quality of input received in the two (or more) languages that they learn. As quantity and quality of input have been shown to be a crucial variable in moderating linguistic performance (for an overview see Unsworth, 2016), it is not clear if the evidence obtained from different bilingual groups is applicable to foreign language learners.

In addition, we use the term ‘foreign language attainment’ to refer to a subset of different foreign language subskills (e.g. pronouncing foreign language sounds, comprehending spoken words, reading and spelling words, etc.). With the aim to capture a representative picture of the foreign language performance of children with poor literacy skills across many subskills, all of the studies reported in this thesis include a broad range of foreign language measures. This is important, because it is possible that children with poor literacy skills struggle in some, but not all foreign language subskills. The measures used in this thesis reflect different aspects of language processing that are required to complete different tasks, as specified in models of language processing. More specifically, we framed our work within Ellis and Young's (1988) model of language processing, because it has been widely used to describe individual differences in native language performance of children with poor literacy skills (Friedmann, Biran & Dotan, 2013; Friedmann & Coltheart, 2016; Kezilas, Kohnen, McKague & Castles, 2014; Kohnen, Nickels, Castles, Friedmann &

McArthur, 2012; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003).[3] In this way, we aimed to identify potential strengths and weaknesses across foreign language subskills that can provide valuable information for teachers and parents and guide the implementation of support strategies. This aspect is specifically addressed in Chapter 4.

Finally, with respect to the term 'individual differences' we refer to differences between the performances of individual children, as well as to differences within one child's performance on different tasks. In this thesis we aimed to not only document individual differences in foreign language attainment of children with poor literacy skills, but also to investigate potential sources that can explain these individual differences. Past research has identified several child-internal characteristics (e.g. native language skills, short term/ working memory capacity, foreign language learning motivation, etc.) that can explain individual differences in foreign language attainment in in unselected populations (for a review see Dörnyei, 2005; Pawlak, 2012; Skehan, 1989). We build on this line of work and extend it to the special case of foreign language attainment in children with poor literacy skills.

As native language skills have been shown to be one of the most important sources of individual differences in foreign language attainment of unselected populations (Sparks & Ganschow, 1991; Sparks, Ganschow, Javorsky, Pohlman & Patton, 1992a, 1992b; Sparks, Ganschow & Pohlman, 1989; Sparks, Patton, Ganschow & Humbach, 2011), we specifically investigated their contribution to explain the variability observed in the foreign language performance of poor readers/spellers. More specifically, we aimed to provide information on the native language deficits that might put individual poor readers/spellers especially at greater risk of experiencing difficulties in equivalent foreign language subskills. For this purpose, we drew on contributions from current models of bilingual processing (i.e. Dijkstra

[3] We describe this model in detail in Chapter 3.

& van Heuven, 2002; Grosjean, 1988, 1997; Kroll & Stewart, 1994) and integrated them into Ellis and Young´s (1988) language processing model, which has been used extensively to identify underlying impairments in children with poor literacy skills. This aspect was specifically addressed in the study reported in Chapter 3.

All the empirical studies presented in this thesis focused on the language pairing of German and English. In the studies reported in Chapter 2 and 3 we assessed native German-speaking children learning English as a foreign language, while children in the study in Chapter 4 were Australian English native speakers receiving German foreign language instruction. German and English are similar in many ways due to their common historical origin as Germanic languages (König & Gast, 2012). There are, for example, many words with the same or similar pronunication, spelling and meaning in both languages that are called cognates (e.g. *fish* in English and *Fisch* in German; König & Gast, 2012). However, German and English also show important differences, for example, with respect to their orthographies. In German successful reading and spelling for most words can be completed by relying on regular grapheme-phoneme correspondences. In contrast, English contains a higher proportion of written words with irregular grapheme-phoneme correspondences (Frith, Wimmer & Landerl, 1998; Landerl, 2017; Landerl, Wimmer & Frith, 1996). These similarities and differences were specifically taken into account in the design of the test materials and the interpretation of the results reported in Chapter 2, 3 and 4.

In summary, the overall aim of this thesis is to contribute towards a better understanding of individual differences in foreign language attainment of children with poor literacy skills. **Chapter 1** sets the stage for the subsequent chapters by systematically reviewing existing work on how successful children/adolescents with poor literacy skills are in learning a foreign language, as compared to children/adolescents with typical literacy skills. To capture a broad range of foreign language skills researched by previous studies, we collected data on 15 different foreign language measures. Moreover, we also explored the influence of moderators related to participant characteristics, foreign language instruction and assessment on the results of the studies that met inclusion criteria. From a methodological point of view, our meta-analyses did not only include overall effects comparing the mean average performance across participant groups, but also provided information on the magnitude of performance variability between poor and typical readers/spellers. This is an innovative meta-analytic procedure that has recently proven useful to determine the magnitude of inter-subject variability in the field of biological evolution and nutrition, and we provide the first application of this to synthesize available evidence on children/adolescents with poor literacy skills (Nakagawa et al., 2015; Senior, Gosby, Lu, Simpson & Raubenheimer, 2016).

Chapter 2 builds on the results of the systematic literature review and asks to what extent individual children with poor literacy skills show a lower attainment than children with typical literacy skills. To address this question, we collected information on eight foreign language subskills in German speaking children with poor and typical literacy skills learning English as a foreign language at school. In line with past research, we first compared group averages between the performance of children with poor and typical literacy skills. However, in a second step we used single case statistics to investigate to what extent group averages reflected the individual performance of each poor reader/speller in this study. These analyses also allowed us to explore if individual children with poor literacy skills show difficulties in different foreign language subskills.

Based on the findings reported in Chapter 1 and 2 (that foreign language attainment in children with poor literacy skills is a highly heterogeneous process), in **Chapter 3** we aimed to identify potential sources of this variability. Therefore, we asked if native language subskills contribute to explaining individual differences in equivalent foreign language subskills in poor readers/spellers. To address this question, we performed a model-based analysis of cross-linguistic interdependence between native and equivalent foreign language subskills in the same poor readers/spellers that we reported on in Chapter 2. More specifically, we integrated contributions from current models of bilingual processing into a well-known language processing model (i.e. Ellis & Young, 1988) that has a longstanding history of being applied in clinical settings to diagnose language deficits across different languages (Friedmann, Biran & Dotan, 2013; Friedmann & Coltheart, 2016; Kezilas, Kohnen, McKague & Castles, 2014; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003). As a secondary aim of this study, we also explored the role of linguistic background, intellectual ability, short term and working memory capacity and foreign language learning motivation as potential sources of individual variability in foreign language attainment of children with poor literacy skills.

In **Chapter 4** we deepen our analysis of individual differences in foreign language attainment of children with poor literacy skills by focusing specifically on foreign language word knowledge. We extend the lexical quality framework (Perfetti & Hart, 2002; Perfetti, 2007, 2017), which has mainly been used to explore native language word knowledge, to investigate how variable the performance of individual children is across different tasks, in which foreign language tasks children show the strongest and the weakest performance and if different psycholinguistic variables influence foreign language word knowledge. To address these questions, we analysed individual performance profiles of ten English speaking children with poor literacy skills who were learning German as a foreign language on six different tasks involving the same 47 German words.

Finally, in the **General Discussion** we establish connections between the findings of each of the above-mentioned studies with the aim of providing a coherent overview of individual differences in foreign language attainment of children with poor literacy skills. Moreover, we discuss theoretical, methodological and practical implications and reflect on the contributions of this thesis, as well as on potential future directions.

General introduction

This thesis is written following a "thesis by publication" format. Therefore, each chapter is presented as an independent manuscript. As each chapter addresses a different aspect of the same overall topic, some parts of the text contain overlapping information. The manuscripts reported in Chapters 1 and 3 are currently under review in two different peer reviewed journals, while the information in Chapter 2 and 4 have not yet been submitted.

References

Afflerbach, P. (2015). *Handbook of Individual Differences in Reading: Reader, Text, and Context*. New York: Routledge.

Baker, C. (1996). *Foundations of Bilingual Education and Bilingualism*. (2.ed.) UK: Multilingual Matters Ltd.

Bialystok, E. (2012). The impact of bilingualism on language and literacy development. In W.C. Ritchie, & T.K. Bhatia (Eds.) *The Handbook of Bilingualism and Multilingualism. Second Edition* (pp. 624-648). London: Blackwell Publishing.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*(01), 69-85. doi: 10.1017/S0142716400010584

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

British Dyslexia Association (2016). Modern Foreign Languages. Is it possible for someone who is dyslexic to learn a foreign language? Available at http://www.bdadyslexia.org.uk/educator/modern-foreign-languages [accessed 20/10/2016].

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197. doi: 10.1017/s1366728902003012

Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate*. New York: Cambridge University Press.

Ellis, A.W. & Young, A.W. (1988) *Human cognitive neuropsychology*. Hove: Erlbaum.

Farukh, A., & Vulchanova, M. (2016). L1, quantity of exposure to L2 and reading disability as factors in L2 oral comprehension and production skills. *Learning and Individual Differences*, *50*, 221-233. doi: 0.1016/j.lindif.2016.07.013

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. In A. Bar-On, & D. Ravid (Eds.), *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistics Perspectives*. Berlin, Boston, MA: De Gruyter Mouton.

Friedmann, N., Biran, M., & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K. K. Grohmann (Eds.), *The Cambridge Handbook of Biolinguistics* (pp. 350-374). Cambridge, UK: Cambridge University Press.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German-and English-speaking children. *Scientific Studies of reading*, *2*(1), 31-54.

Genesee, F. (2015). Myths About Early Childhood Bilingualism. *Canadian Psychology* 56(1), 6-15. doi: 10.1037/a0038599

Germanò, E., Gagliano, A., & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental neuropsychology*, *35*(5), 475-493. doi: 10.1080/87565641.2010.494748

Ghonsooly, B. & Javadian, M. (2010). An Examination of Developmental Dyslexia among Iranian EFL Second Graders. *International Journal of Language Studies*, 4(4*),* 333-356.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. Language and Cognitive Processes, 3, 233–274. doi: 10.1080/01690968808402089

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A.M.B. de Groot, J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, (pp. 225-254). NJ: Erlbaum.

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

Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of Psycholinguistic Research*, 34(6), 603-618. doi: 10.1177/0022219409345018.

Hussien, A. M. (2014). The effect of learning English (FL) on learning of Arabic literacy (NL) in the primary school. *International Education Studies*, 7(3), 88-98. Retrieved from http://www.ccsenet.org/journal/index.php/ies/article/viewFile/33294/19765.

International Dyslexia Association (2012). Definition of dyslexia. Available at https://dyslexiaida.org/definition-of-dyslexia/ [accessed 20/10/2016].

Kezilas, Y., Kohnen, S., McKague, M., & Castles, A. (2014). The locus of impairment in English developmental letter position dyslexia. *Frontiers in human neuroscience*, *8*, 356. doi: 10.3389/fnhum.2014.00356

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime' becomes 'smile': Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681-3692. doi:10.1016/j.neuropsychologia.2012.07.016

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex*. doi: 10.1016/j.cortex.2018.04.008

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false‐belief reasoning. *Developmental science*, *12*(1), 48-54. doi:10.1080/19404158.2015.1023209

König, E. & Gast, V. (2012). *Understanding English-German Contrasts*. Berlin: Erich Schmidt Verlag.

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

Landerl, K. (2017). Learning to Read German. In L. Verhoeven & C. Perfetti (2017). *Learning to Read across Languages and Writing Systems* (pp. 299-322). United Kingdom: Cambridge University Press.

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*(3), 287-294. doi: 10.1111/j.1469-7610.2009.02164.x

Landerl, K., Wimmer, H., & Frith, U. (1996). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*(3), 315-334. doi: 10.1016/S0010-0277(97)00005-X.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological methods*, *7*(1), 19-40. doi: 10.1037//1082-989X.7.1.19

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “specifics” of specific reading disability and specific language impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(7), 869-874.

McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E. & Castles, A. (2013). Getting the grips with the heterogeneity of developmental dyslexia. *Cognitive Neuropsychology* 30(1), 1-24. doi: 10.1080/02643294.2013.784192

Miller-Guron, L., & Lundberg, I. (2000). Dyslexia and second language reading: A second bite at the apple?. *Reading and Writing*, *12*(1-2), 41-61. doi: 10.1023/A:1008009703641

Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*(5), 359-382. doi: 10.1080/10888430903162878

Moll, K., Göbel, S. M., & Snowling, M. J. (2015). Basic number processing in children with specific learning disorders: Comorbidity of reading and mathematics disorders. *Child neuropsychology*, *21*(3), 399-417. doi: 10.1080/09297049.2014.899570

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143-152. doi: 10.1111/2041-210X.12309

Palladino, P., Bellagamba, I., Ferrari, M., & Cornoldi, C. (2013). Italian children with dyslexia are also poor in reading English words, but accurate in reading English pseudowords. *Dyslexia*, *19*(3), 165-177. doi: 10.1002/dys.1456

Pawlak, M. (2012). New Perspectives on Individual Differences in Language Learning and Teaching. Berlin, Heidelberg: Springer.

Peer, L. & Reid, G. (Eds.) (2014). *Multilingualism, literacy & dyslexia*. UK: Taylor & Francis Group.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. reitsma (Eds.), *Precursors of functional literacy* (pp.189-213). Amsterdam: John Benjamins. doi: 10.1075/swll.11.14per.

Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. doi:10.1080/10888430701530730.

Perfetti, C.A. (2017). Lexical quality revisited. In E. Segers & P. van den Broek. *Developmental Perspectives in Written Language and Literacy* (pp. 51-67). Amsterdam: John Benjamins Publishing Company.

Schulte-Körne, G. (2010). The prevention, diagnosis, and treatment of dyslexia. *Deutsches Ärzteblatt International*, *107*(41), 718-727. doi: 10.3238/arztebl.2010.0718

Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, medicine, and public health*, *2016*(1), 244-255. doi: 10.1093/emph/eow020.

Serry, T. A., & Hammond, L. (2015). What's in a word? Australian experts' knowledge, views and experiences using the term dyslexia. *Australian Journal of Learning Difficulties*, *20*(2), 143-161. doi: 10.1080/19404158.2015.1089916

Siegel, L. S. (2018). Learning disabilities: The roads we have travelled and the path to the future. In R. Sternberg & L Spear-Swerling (Eds.) *Perspectives on learning disabilities: biological, cognitive, contextual* (pp. 159-175). New York: Routledge.

Siegel, L.S. (1988a). Definitional and theoretical issues and research on learning disabilities. *Journal of Learning Disabilities*, 21, 264-266.

Siegel, L.S. (2007). Perspectives on dyslexia. *Paediatrics & Child Health*, 11, 581-588.

Skehan, P. (1989). *Individual Differences in Second Language Learn*ing. London: Edward Arnold.

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216. doi: 10.1016/j.cognition.2017.06.024

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreign Language Annals*, 49(2), 252-270. doi:10.1111/flan.12196.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal*, 75(1), 3-16. doi:10.2307/329830.

Sparks, R. L., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992a). Identifying native language deficits in high- and low-risk foreign language learners in high school. *Foreign Language Annals*, 25(5), 403.

Sparks, R. L., Humbach, N., Patton, J. O. N., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, *95*(2), 253-273. doi:

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia*, 56(1), 129-160. doi: 10.1007/s11881-006-0006-2.

Sparks, R., Ganschow, L., & Pohlman, J. (1989). Linguistic coding deficits in foreign language learners. *Annals of Dyslexia*, 39(1), 177-195. doi:10.1007/BF02656908.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992b). Test Comparisons among Students Identified as High-Risk, Low-Risk, and Learning Disabled in High School Foreign Language Courses. *The Modern Language Journal*, 76(2), 142-159. doi:10.2307/329768.

Stadie, N., & van de Vijver, R. (2003). A linguistic and neuropsychological approach to remediation in a German case of developmental dysgraphia. *Annals of Dyslexia*, *53*(1), 280-299.

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.) *Bilingualism across the lifespan: Factors moderating language proficiency* (p.103-122). Washington, DC: Mouton de Gruyter, American Psychological Association.

van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of Learning Disabilities*, 39(1), 74-90. doi:10.1177/00222194060390010701.

Wight, M.C.S. (2015). Students With Learning Disabilities in the Foreign Language Learning Environment and the Practice of Exemption. *Foreign language Annals*, 48(1), 39-55. doi: 10.1111/flan.12122.

Zoccolotti, P., & Friedmann, N. (2010). From dyslexia to dyslexias, from dysgraphia to dysgraphias, from a cause to causes: a look at current research on developmental dyslexia and dysgraphia. *Cortex*, *46*(10), 1211-1215.

Chapter 1

Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis

This chapter is currently under review as:

von Hagen, A., Kohnen, S. & Stadie, N. (under review). Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis.

Abstract

This systematic review investigated how successful children/adolescents with poor literacy skills are compared to their peers with typical literacy skills in learning a foreign language. Moreover, we explored whether specific characteristics related to participants, foreign language instruction and assessment moderated scores on foreign language tests in this population. Overall, 16 studies with a total of 968 participants (poor reader/spellers: n = 404; control participants: n = 564) met eligibility criteria. Available data allowed for meta-analyses on 10 different measures of foreign language attainment. In addition to standard mean differences (*SMDs*), we computed natural logarithms of the ratio of coefficients of variation (*CVRs*) to capture individual variability between participants groups. Significant between-study heterogeneity, which could not be explained by moderator analyses, limited the interpretation of results. Although children/adolescents with poor literacy skills on average showed lower scores on foreign language spoken word production, phonological awareness, letter knowledge, and reading comprehension measures, their performance varied significantly more than that of control participants. Thus, it remains unclear to what extent group differences between the foreign language scores of children/adolescents with poor and typical literacy skills are representative of individual poor readers/spellers. Taken together, our results indicate that foreign language skills in children/adolescents with poor literacy skills are highly variable. We discuss limitations of past research that can guide future steps towards a better understanding of individual differences in foreign language attainment of children/adolescents with poor literacy skills.

Keywords: poor literacy, dyslexia, foreign language, bilingualism, meta-analysis

Public significance statement

This systematic review shows that the success children/adolescents with poor literacy skills achieve in learning a foreign language as compared to their peers with typical literacy skills is highly variable. Research to date cannot explain why some poor readers/spellers show a lower foreign language attainment, while others are just as successful as their peers with typical literacy skills. This review highlights the need for future research in order to reach a better understanding of this issue so that teachers and parents can make evidence-based educational decisions for poor readers/spellers.

Introduction

While research has investigated the native language abilities of children/adolescents with poor literacy skills quite extensively, less attention has been paid to their success in learning a foreign language in formal educational settings. Relatives, teachers and allied health professionals often assume that the difficulties children/adolescents with poor literacy skills experience in their native language will transfer to the new language being learned. For example, the British Dyslexia Association highlights that "dyslexic pupils learning a foreign language have to be aware that it may be a longer process for them than for others" (British Dyslexia Association, 2016). However, they advocate that each student should have the opportunity to receive foreign language instruction. Sparks (2016) summarizes that the educational field sustains the belief that students identified with learning difficulties (including poor literacy skills) will struggle in learning a foreign language.

As a consequence, children/adolescents with poor literacy skills receive less support and are even exempted by law from foreign language instruction in many countries. One example of many is an Italian law, allowing children/adolescents with poor literacy skills to be completely excused from foreign language learning (see Palladino, Bellagamba, Ferrari & Cornoldi, 2013). Researchers in the field of foreign language learning, express their concern about these policies, as to date, the evidence regarding foreign language difficulties in poor readers/spellers is scarce. Wight (2015) suggests that the policies and practices of exempting students from foreign language study demonstrate that they are often discharged "(1) based on personal beliefs and preferences rather than on the basis of a carefully considered consensus of inclusion, and (2) in the absence of actual data about the potential successes of students with special needs" (pp.41-42, Wight, 2015).

Although these policies aim to protect children/adolescents with poor literacy skills from experiencing failure, they also impede those students to gain cognitive and professional advantages associated with foreign language learning (e.g. advantage on inhibitory control -

Bialystok & Majumber, 1998; advantage in theory of mind development – Kovacs, 2009; advantage on metalinguistic knowledge – Bialystok, 2012). Moreover, access to cultural diversity remains limited without being able to speak an additional language. Despite having a profound impact on students' future opportunities, these decisions are -to date- not based on a systematic evaluation of the existing evidence.

Therefore, a systematic review that investigates how successful children/adolescents with poor literacy skills are in learning a foreign language is urgently needed. Insights into past research on this topic should support families, teachers and specialists working with children/adolescents with poor literacy skills in making informed decisions about foreign language instruction. In the following sections we define the main elements of our systematic review according to the PICO acronym (Population, Intervention, Comparison, Outcome; O'Connor, Green & Higgins, 2008). This includes information on the population investigated in this review (children/adolescents with poor and typical literacy skills), the intervention received by participants (foreign language instruction) and the outcomes that were measured (foreign language attainment). Furthermore, in each section we detail potential moderators that may impact foreign language attainment in children/adolescents with poor literacy skills.

Children/adolescents with poor literacy skills

Worldwide, a significant proportion of children/adolescents present with poor literacy skills that cannot be explained by medical, emotional or neurological difficulties or insufficient literacy instruction. Prevalence rates of children/adolescents with poor literacy skills range from 3.1% to 17.5% across languages (e.g. 3.1 - 3.2% in Italian, Barbiero et al., 2012; 5% in German, Müller et al., 2014; 17.5% in American English, Shaywitz, Morris & Shaywitz, 2008). Many different terms have been used to describe poor literacy skills: e.g. developmental dyslexia and/or dysgraphia, specific reading and/or spelling difficulty, reading and/or spelling impairment, deficit or disability (Elliott & Grigorenko, 2014; Siegel, 1988a, 2007). In the present review we use the term 'children/adolescents with poor literacy skills' or 'poor readers/spellers' to refer to students experiencing difficulties in reading and/or spelling.

Poor literacy, in the context of this review, encompasses reading and spelling difficulties. With respect to reading difficulties, we included inaccurate or slow reading of words, nonwords, sentences or texts (IDA, 2012). While some poor readers/spellers may additionally struggle with reading comprehension tasks, students solely facing reading comprehension difficulties and no other reading deficits (e.g. inaccurate or slow word or nonword reading), have not been included in this review into the group of children/adolescents with poor literacy skills. This is mainly due to the fact that reading comprehension difficulties have been shown to be associated with oral language deficits (e.g. poor vocabulary knowledge, poor sentence comprehension skills), especially when they occur without additional difficulties with reading accuracy and/or fluency (e.g., Oakhill, Cain, & Bryant, 2003). Concerning spelling difficulties, children/adolescents with poor literacy skills may struggle in spelling words or nonwords either in dictation tasks and/or in spontaneous text production (Kohnen, Colenbrander, Krajenbrink & Nickels, 2015). In this review 'children/adolescents with poor literacy skills' have any combination of the abovementioned reading and spelling difficulties.

Past research has shown that children/adolescents with poor literacy skills have very heterogeneous performance patterns in their native language (Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013; Moll & Landerl, 2009). This heterogeneity may extend to foreign language skills such that only some, but not all poor readers/spellers show a lower attainment than their peers with typical literacy skills. Thus, potential moderators related to participant characteristics are likely to influence the magnitude of performance differences between children/adolescents with poor and typical literacy skills. Selective impairments in native language profiles of poor readers/spellers might impact foreign language attainment in different ways. Some of the profiles that have been described are for example poor readers/spellers that only show difficulties in written native language skills, while others also show impaired oral language abilities (Bishop & Snowling, 2004). Within the written language domain, at least three profiles have been distinguished: poor readers/good spellers, good readers/poor spellers and poor readers/poor spellers (Moll & Landerl, 2009). Furthermore, different types of reading and spelling deficits have been reported (e.g. Coltheart & Kohnen, 2012; Friedmann & Coltheart, 2016; Hanley, 2017; McArthur et al., 2013;). Some students mainly struggle in converting letters to sounds (i.e. sublexical impairment; e.g. Friedmann & Coltheart, 2016; McArthur et al., 2013). Others have difficulties in recognizing written words, leading to inaccurate or slow word reading (i.e. lexical impairment (e.g. Friedmann & Lukov, 2008; Kohnen, Nickels, Geigis, Coltheart, McArthur, & Castles, 2018; Sotiropoulos & Hanley, 2017). Moreover, some children show deficits in processing letter order leading to excessive anagram mistakes (e.g. Friedmann & Rahanim, 2007; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012). In addition, deficits in other processes subserving reading performance, such as recognizing letters, ordering letters and moving letters between words, have also been observed (e.g. Friedmann & Coltheart, 2016). It is currently unknown if all of the above-mentioned selective impairments of poor readers/spellers impact foreign language attainment in the same way.

Another moderator that might contribute to variable foreign language performance in children/adolescents with poor literacy skills is the diversity of linguistic backgrounds. Several studies have reported higher foreign language attainment in bilingual as compared to monolingual students with typical literacy skills. For example, Nair, Biedermann and Nickels (2015) found a significantly better performance in early and late bilinguals as compared to monolinguals in a novel-word-learning task. Similarly, Tremblay and Sabourin (2012) reported significantly higher speech perception abilities in multi- and bilinguals as compared to monolinguals.

Foreign language instruction

Foreign language instruction refers to classroom-based interaction between students and teachers with the aim of improving the proficiency of a child/adolescent to communicate in a language other than his or her native language. More specifically, this means that the child/adolescent has no additional exposure to the foreign language, other than limited access for example to music, films and computer games. This is different to other situations in which children learn an additional language, for example through the exposure to more than one language from birth. Also, additional language learning through immersion in an environment in which this language is dominant (e.g. when a family moves to a new country, children of immigrant families or so-called 'English-language learners') is not meant by the term 'foreign language learning' in this review.

The manner in which a foreign language is being instructed also influences its attainment (Saito & Hanzawa, 2016). Many different instruction types have been described that can broadly be categorised into ‘language-based’ versus ‘content-based’ approaches (Wright, 2013). Language-based approaches focus on the instruction of the foreign language itself, whereas content-based approaches make use of different school subjects such as music or history to teach the foreign language. Similarly, some authors differentiate between foreign language instruction and immersion emphasizing the quantity and quality of foreign language input (Wright, 2013). Therefore, duration and frequency of foreign language input might be potential moderatos of foreign language attainment in children/adolescents with poor literacy skills. Lastly, onset age of foreign language instruction has been shown to influence foreign language skills in children/adolescents with typical literacy (e.g. Bialystok, 1997; DeKeyser, Alfi-Shabtay & Ravid, 2010; Friederici, Steinhauer & Pfeifer, 2002; Johnson & Newport, 1989).

Language pairing between native and foreign language has also been shown to moderate foreign language attainment (Connor, 1996; Odlin, 1989). Structural similarities or differences between the native and foreign language, for example between Indo-European and non-Indo-European languages, have been shown to either facilitate or impede the acquisition of a foreign language (Connor, 1996; Melby-Lervag & Lervag, 2014; Odlin, 1989). In addition, Bialystok, Luk and Kwan (2005) pointed out that the orthographic similarity between two writing systems (e.g., alphabetic versus ideographic writing systems) was able to explain the extent to which bilingual children were able to positively transfer literacy skills across languages. Moreover, within alphabetic writing systems, it seems that the regularity with which a grapheme is mapped onto a phoneme is an important moderator of literacy performance (e.g. Seymour, Aro & Erskine, 2003; Ziegler et al, 2010). This is especially relevant in children/adolescents with poor literacy skills. Although students with poor literacy skills have been identified in many different languages (e.g., Frost, 2012; Ziegler, Perry, Ma-Wyatt, Ladner & Schulte-Körne, 2003), some performance patterns seem to differ across languages (Goulandris, 2003; Landerl, Wimmer & Frith, 1996; Moll et al., 2014).

Foreign language attainment

The age at which foreign language abilities of poor readers/spellers are assessed could be another potential moderator. Indeed, Bialystok (1997) highlighted that older learners rely on wider previous knowledge than younger ones and are therefore able to include new information into already existing conceptual categories. In contrast, younger learners tend to create new categories for the input they receive, which sometimes involves a longer learning process. Similarly, DeKeyser (2000) reported that younger learners rely to a greater extent on implicit mechanisms that may no longer be available to older learners. Older learners depend much more on explicit learning mechanisms. Both types of learning mechanisms have been shown to be beneficial in developing different foreign language subskills. For instance, it appears that speech production relies to a greater extent on implicit learning, while grammatical knowledge is acquired faster through explicit teaching (DeKeyser, 2000). Thus, the age at which a foreign language assessment is conducted might impact the magnitude of group differences between children/adolescents with poor and typical literacy skills.

Foreign language attainment also involves mastering distinct subskills, such as for example discriminating foreign language speech sounds, comprehending spoken words or reading and spelling words. It may therefore be possible that children/adolescents with poor literacy skills only show a lower performance in some, but not all foreign language subskills. A detailed investigation of existing research on different foreign language subskills of poor readers/spellers can shed light on this issue. However, as different research traditions in the foreign language learning literature have used different labels to describe these subskills, it is sometimes difficult to reconcile classification systems. For example, some authors distinguish between oral and written language, while others contrast receptive and expressive modalities (Nation, 2013). Again, others segregate tasks focusing on pre-lexical, lexical and non-lexical processing mechanisms (de Bot, 1992; de Bot, Paribakht & Wesche, 1997). For the purpose of this review, we consider a classification based on the tasks used to measure foreign language attainment (e.g. picture naming, rhyme detection, lexical decision, etc.) as the most appropriate one.

The current review

In the current review we systematically analysed available evidence aiming to address the following two research questions:

1. How successful are children/adolescents with poor literacy skills in learning a foreign language, as compared to children/adolescents with typical literacy skills? To capture a complete picture of foreign language skills in poor readers/spellers, we compared existing information on a broad range of oral and written foreign language outcome measures.

2. Is successful foreign language attainment in children/adolescents with poor literacy skills influenced by moderators such as participant characteristics, foreign language instruction and foreign language assessment? Information on a broad range of moderators was

collected to investigate under which conditions children/adolescents with poor literacy skills are just as successful as their peers with typical literacy skills in learning a foreign language.

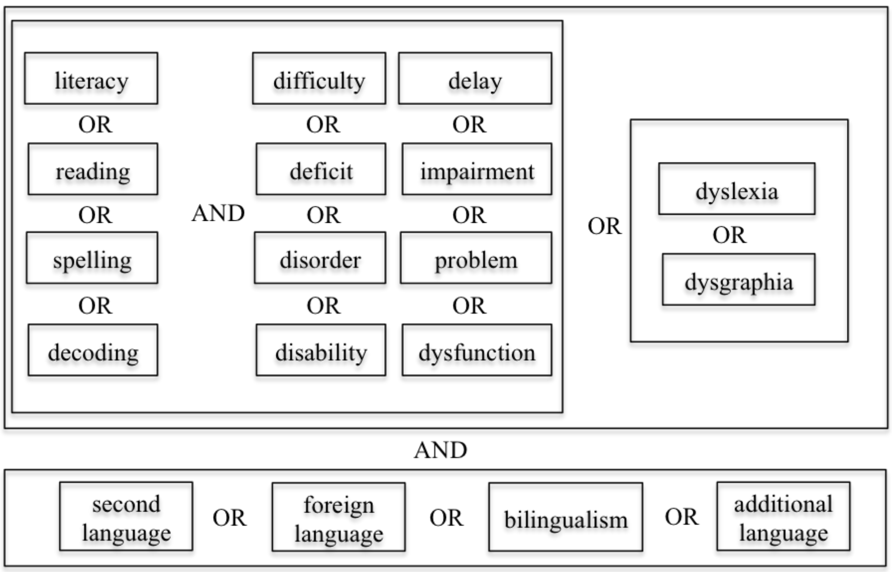
Method

The procedures used in the current review were pre-defined in a protocol registered in the *International Prospective Register of Systematic Reviews* PROSPERO. This document is available at https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=69980. Differences between the protocol and this final report are detailed in supplemental materials.

Literature search

Search strategy. All the search terms and combinations shown in Figure 1 were entered into the databases detailed below in the section of information sources on the 10th and 26th of February of 2017. This search strategy was adapted to the requirements of the different databases we used (see supplemental materials). No language or date limits were applied.

Figure 1. Search terms and strategy



Titles and abstracts of all identified references were screened. Following this, full texts of references that met inclusion criteria (see section eligibility of studies) were reviewed. Next, we screened the reference lists of the studies that met inclusion criteria in order to detect other studies that were not found with the above mentioned search terms (also see

Francis, Caruana, Hudson & McArthur, 2016). Lastly, we entered the titles of the included full text reports into the 'cited by' function of GoogleScholar and screened titles and abstracts of the studies that cited each of the included studies.

Information sources. We searched the following databases: Ovid databases (PsycINFO, PsycARTICLES, MEDLINE, EMBASE), Wiley Database, PubMed, ProQuest (ERIC, ProQuest dissertations and Linguistics and Language Behaviour Abstracts (LLBA)) and Web of Science. Furthermore, PsycExtra and GoogleScholar were used to identify grey literature. In addition, the following journals were hand searched: *International Journal of Bilingual Education and Bilingualism, Bilingualism: Language and Cognition, Second Language Research, TESOL Quarterly,* and *International Journal of Bilingualism*. We also planned to contact authors of more than three independent studies that met inclusion criteria per e-mail and ask for unpublished material.

Eligibility of studies

We summarized the selection criteria in nine 'signalling questions' that guided the screening procedure of all the study reports (see Appendix 1). A study report was removed from the review when at least one response indicated exclusion.

Participants. In order to meet inclusion criteria, a study had to report on (a) children/adolescents with poor literacy skills and (b) children/adolescents with typical literacy skills. Children/adolescents with poor literacy skills had to present below average scores (i.e. either one standard deviation, one year or grade below the expected level) in either a reading or spelling task or in both (Elliott & Grigorenko, 2014). Studies in which participants were included based on self- or teacher reports were excluded (Snowling, Duff, Petrou & Schiffeldrin, 2011). Furthermore, we only incorporated studies assessing participants from their first year of formal schooling up to the last year of secondary education, whereas studies investigating students in post-secondary education were excluded.

Concerning the type of reading difficulty, studies with participants showing inaccurate or slow performance in reading either words, nonwords, sentences or texts were included. In contrast, we eliminated studies in which only reading comprehension was assessed (Oakhill, Cain & Bryant, 2003). With respect to spelling difficulties, participants had to reveal below average word or nonword spelling (i.e. either one standard deviation, one year or grade below the expected level) in order to meet inclusion criteria for the present review. These difficulties could have been assessed through tasks such as word, nonword, sentence or text dictation or through spontaneous text productions.

All studies included in the meta-analysis compared foreign language performance of poor readers and spellers with control participants demonstrating typical literacy skills. Once again, reading or spelling tests had to be used to confirm typical literacy skills. Control participants had to score within the expected age-group, grade or less than one standard deviation below the expected level (Elliott & Grigorenko, 2014).

Types of foreign language instruction. All types of classroom based foreign language instruction (i.e. language vs. content-based approaches; instruction vs. immersion approaches) were considered. However, the foreign language being instructed could not be classified as a national language of the country in which the study took place. Thus, all the studies stating that participants had additional access to the foreign language, other than limited access for example to music, films or travel experiences, were excluded. Examples of this are studies focusing on the foreign language learning abilities of heritage speakers (i.e. children/adolescents that are exposed to a minority language at home).

Foreign language outcome measures. We included studies in which the foreign language was assessed in either the oral and/or written modality and removed studies that only investigated foreign language learning motivation. Foreign language skills could have been measured with self-developed or standardized language tests.

Types of studies. Only group comparison studies (as opposed to case studies or case series) were included. A quantitative comparison of both participant groups on a foreign language outcome measure had to be present. In addition to the above-mentioned eligibility criteria, we only excluded data reported within doctoral dissertations if the same information was published in a peer-reviewed article.

Selection process

First, the titles and abstracts of the studies were imported into Covidence (www.covidence.org) and duplicated references were automatically deleted. Further studies that were clearly identified as duplicates were dismissed. Three reviewers (the authors) each screened two thirds of all of all of the abstracts so that across the three reviewers all titles and abstracts were screened twice. In cases of disagreements between two reviewers, the third reviewer was asked to resolve the conflict through a third judgment.

Second, we downloaded the full version of the included studies, saved them in Mendeley and imported them to Covidence. Once again, the three reviewers (the authors) each read two thirds of all of the papers' introduction, methods and results so that across the three reviewers every paper was read twice. If the available information fulfilled the eligibility criteria for the present review, the study was included. If a study was not eligible, we registered the reasons for this in Covidence. Once again, divergences between reviewers were resolved by the third reviewer.

Risk of bias assessment

Two reviewers independently assessed the risk of bias of each included study by applying an adaptation of the ROBINS-I rating scale (Sterne et al., 2016) (see supplemental materials). This scale focuses on judging the quality of each study within the following domains: (a) confounding, (b) selection of participants into the study, (c) classification of interventions, (d) deviations from intended interventions, (e) missing data, (f) measurement of outcomes, and (g) selection of the reported result. For the last three domains, we did not only assess each study, but also each outcome measure within every study. We used signalling questions to facilitate a domain-level risk of bias judgment (see supplemental materials).

The risk of bias was labelled low, moderate, serious, critical or no information. Again, judgments from two reviewers were juxtaposed and in case of disagreements, decisions were discussed between all three reviewers, leading to a re-evaluation of the risk of bias in some cases. The entire process was completed in Covidence. According to Sterne et al. (2016), most non-randomized interventions studies will at least present an overall moderate risk of bias. For the present review, all studies with serious or critical risk of bias were excluded.

Data collection process

In order to extract the relevant data from included studies, we customized a data extraction form in Covidence (see supplemental materials). The first author of this review completed the data extraction form for each of the included studies. The other two reviewers double-checked this step of the process individually by comparing the data entries with the original studies. Any incongruities were resolved by returning to the original data in the study. In cases of missing data, the corresponding authors of studies were contacted and if this information was not available, the study (or measure) was excluded from the review.

Foreign language outcome measures. We collected data on 15 oral and written foreign language tasks (see Table 1).

Table 1

Data extraction categories for foreign language outcome measures

FL[a] outcome measure	Examples of tasks
Discrimination of speech sounds	Auditory discrimination of phonemes, syllables, words or nonwords
Production of speech sounds	Repetition of phonemes, syllables, words or nonwords
Receptive vocabulary knowledge	Spoken word-picture matching
Spoken word production	Picture naming
	Serial rapid naming
	Semantic fluency
Sentence comprehension	Spoken sentence-picture matching
	Grammaticality judgments
Sentence production	Sentence elicitation
	Picture description
	Conversation
Short term memory	Digit repetition
Phonological awareness	Rhyme detection
	Spoonerisms
	Initial and final phoneme deletion
Letter knowledge	Speeded and unspeeded letter naming
Word reading	Speeded and unspeeded word reading
	Speeded sentence and text reading
Nonword reading	Speeded and unspeeded nonword reading
Orthographic knowledge	Visual lexical decision
Reading comprehension	Text reading and multiple choice questions
	Open questions
	Cloze
Spelling	Word, nonword and sentence spelling to dictation
	Word or nonword copy
	Written story telling
Translation	Translation of words, sentences or texts

Note. [a]FL = Foreign language

Similar tasks were grouped together under broader terms (e.g. auditory discrimination of phonemes and syllables as discrimination of speech sounds). Included studies presented data on foreign language outcome measures as continuous data. Therefore, we extracted the mean, standard deviation and number of participants of the group of children/adolescents with poor literacy skills and the control group for each outcome measure reported in each study. This information was available for all included studies, except for van Viersen et al. (2017)

who reported Bayesian descriptive data. However, the authors kindly provided the original raw scores for each relevant measure.

Moderators. In relation to our second research question, we also gathered information on 11 moderators related to participant characteristics, foreign language instruction and foreign language assessment (see Table 2).

With respect to participant characteristics, first, data on participants' native language profiles was extracted. We distinguished between information on (a) oral and written language deficits, (b) reading and spelling deficits and (c) reading deficit subtype. Studies reporting information on these moderators were allocated to the following subgroups, respectively: (a) participants with poor oral and written native language or average oral and poor written language, (b) poor readers/good spellers, good readers/poor spellers or poor readers/poor spellers and (c) participants with sublexical, lexical, mixed or other reading deficits. Second, in relation to the participants' linguistic background, we distinguished between studies including only monolingual, bilingual or both monolingual and bilingual participants[4].

---

[4] We planned to allocate studies including pluri-/multilingual participants to the bilingual participant group.

Table 2

Data extraction categories for moderators

Moderators	Subgroups
Participant characteristics	
NL[a] profile	
Oral/written language deficits[b]	Poor oral NL and poor written NL
	Average oral NL and poor written NL
Reading/spelling deficits[c]	Poor readers/good spellers
	Good readers/poor spellers
	Poor readers/poor spellers
Reading deficit subtype[d]	Sublexical reading deficit
	Lexical reading deficit
	Mixed reading deficit
	Other reading deficits
Linguistic background	Monolingual
	Bilingual
	Monolingual and bilingual
FL[e] instruction	
Frequency of FL classes	Less than 2 classes per week
	Between 2-4 classes per week
	More than 4 classes per week
Duration of FL classes	Less than 30 minutes
	Between 30-60 minutes
	More than 60 minutes
Language pairing between NL and FL	
Structural differences[f]	NL Indo-European/ FL Indo-European
	NL Indo-European/ FL non-Indo-European
	NL non-Indo-European/ FL Indo-European
	NL non-Indo-European/ FL non-Indo-European
Writing system differences[d]	Alphabetic NL/ alphabetic FL
	Alphabetic NL/ ideographic FL
	Ideographic NL/ alphabetic FL
	Ideographic NL/ ideographic FL
Orthographic regularity[e]	Regular NL/ regular FL
	Regular NL/ irregular FL
	Irregular NL/ regular FL
	Irregular NL/ irregular FL
Onset age of FL instruction[f]	Early childhood: onset age before 6 years
	Late childhood: onset age from 6-11 years
	Adolescence: onset age from 12-17 years
	Early adulthood: onset age from 18 years onwards
FL assessment	
Age of FL assessment[f]	Early childhood: before 6 years of age
	Late childhood: from 6-11 years of age
	Adolescence: from 12-17 years of age
	Early adulthood: from 18 years of age onwards

Note. [a]NL = Native language; [b]Bishop & Snowling (2004); [c]Moll & Landerl (2009); [d]Friedmann & Coltheart (2016); [e]FL = Foreign language; [f]Melby-Lervag & Lervag (2014); [g]Only categorized within alphabetic writing systems; Seymour et al. (2003); Ziegler et al. (2010); [h] Abrahamson & Hyltenstam (2009).

For foreign language instruction, we extracted data on (a) frequency and (b) duration of foreign language classes, (c) language pairing between native and foreign language and (d) age at onset of foreign language instruction. First, we allocated studies to a subgroup of less than two, between two and four or more than four classes per week. Second, the duration of foreign language classes was categorized into: less than 30 minutes, between 30 and 60 minutes and more than 60 minutes per class. Third, regarding language pairing between native and foreign language, we registered information on (a) structural differences (Indo-European or non Indo-European languages), (b) differences between writing systems (alphabetic or ideographic writing systems[5]) and (c) orthographic regularity in the cases of alphabetic writing systems (regular or irregular). Finally, following Abrahamson and Hyltenstam (2009), we distinguished between four onset ages for foreign language instruction: early childhood (onset age before 6 years), late childhood (onset age from 6 to 11 years), adolescence (onset age from 12 to 17 years) and early adulthood (onset age from 18 years onwards).

With respect to foreign language assessment, we gathered information on the age of participants at the time of the foreign language assessment. Studies were assigned to early or late childhood, adolescence or early adulthood subgroups (see above; Abrahamson & Hyltenstam, 2009).

Data synthesis

Separate meta-analyses were planned for the 15 foreign language outcome measures (see Table 1). However, analyses for each foreign language outcome measure were only completed if information from at least two studies was available (Borenstein et al., 2009). Following common meta-analytic procedures, we used standard mean differences (*SMDs)* with Hedges correction *g* for small sample sizes (Borenstein, Hedges, Higgins & Rothstein,

[5] In this review we only distinguished between alphabetic and ideographic writing systems, although other types exist (see Sampson, 2015).

2009) as the unit of analysis. This allowed us to compare the average foreign language performance between children/adolescents with poor and typical literacy skills.

However, we were concerned that this information might not be representative of the performance of individual children/adolescents with poor literacy skills. Based on the heterogeneous performance of poor readers/spellers documented in past research, the extent to which group averages capture individual performances of children/adolescents with poor literacy skills might be very variable. Results from meta-analyses solely based on standard mean differences (*SMD*s) might therefore have limited potential to guide practical implications for individual children/adolescents with poor literacy skills. To address this limitation, we also computed a second overall effect focusing on the variability across participant groups (the natural logarithm of the ratio between the coefficients of variation of both participant groups - *CVR;* Nakagawa et al., 2015). This allowed us to compare the magnitude of performance variability between poor and typical readers/spellers' foreign performance. Such meta-analytic procedures have recently proven useful to determine the magnitude of inter-subject variability in the field of biological evolution and nutrition (Nakagawa et al., 2015; Senior, Gosby, Lu, Simpson & Raubenheimer, 2016). To our knowledge, this procedure has not been used to synthesize available evidence on children/adolescents with poor literacy skills so far. Nonetheless, given the well-documented heterogeneity of this population (Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013; Moll & Landerl, 2009), adopting this procedure seems justified.

Both types of effect sizes (*SMDs* and *CVRs*) were derived from the mean (*M*), standard deviation (*SD*) and number (*n*) of participants for each foreign language task. Some studies reported information on more than one group of children/adolescents with poor literacy skills or more than one control group. In those cases, we combined the *M*, *SD* and *n* of both groups or excluded one of the groups (Borenstein et al., 2009). Reasons for relevant decisions are detailed in the results section.

Furthermore, many studies used more than one task for the same outcome measure (e.g. phoneme deletion and substitution tasks to assess phonological awareness). In these cases, *SMDs* and *CVRs* were computed separately for each task and subsequently, values were aggregated for each outcome measure (Borenstein et al., 2009). Such aggregation methods take into account the correlation between the aggregated tasks. However, as this information was not available for many studies, we assumed a large correlation of $r = .50$ (Cohen, 1988) based on the similarity of the tasks being aggregated (Borenstein et al., 2009). The same procedure was followed for longitudinal studies reporting more than one data-point per outcome measure (Borenstein et al., 2009).

Before aggregating *SMDs*, we ensured that a negative difference indicated that the control group performed better than the group of children/adolescents with poor literacy skills for all comparisons. If measures were based on the occurrence of errors (instead of accuracy rates), the sign of the *SMDs* was reversed (Borenstein et al., 2009). Likewise, before aggregating *CVRs* we ensured that a negative difference for all comparisons indicated higher performance variability in the group of children/adolescents with poor literacy skills than in the control group.

Based on all potential moderators that could influence foreign language attainment in children/adolescents with poor literacy skills, we expected to find significant heterogeneity between studies. Therefore, we decided a priori to use random effects modelling to consider the study inverse variance and the between-study variance. We used Cochran´s Q statistic with a significance level of $p < .05$ to determine the presence of heterogeneity among effect sizes. Furthermore, to quantify heterogeneity, we calculated τ^2 and I^2, as an index of the variation between study effect sizes. We followed Higgins, Thompson, Deeks and Altman´s (2003) guidelines and considered I^2 values around 25%, 50% and 75% as low, moderate and high heterogeneity, respectively. To measure the overall effect, we used Z statistics with a Bonferroni corrected significance level according to the number of comparisons that we were performing (Borenstein et al., 2008). Overall effects were only interpreted in the absence of significant heterogeneity between study effect sizes (*Q* statistic $p > .05$).

Moderator analyses. In order to explore the impact of specific moderators on the foreign language attainment of children/adolescents with poor literacy skills, we planned to compute separate analyses for the 11 moderators on which we collected information (see Table 2). However, in line with Littell, Corcoran and Pillai (2008), moderator analyses were only computed if data was available from at least 10 studies. We completed separate random mixed modelling meta-analyses for each moderator subgroup using the *metafor* package in *R* (Viechtbauer, 2010). Finally, we used a *Z*-test with an adjusted significance level according to the number of comparisons computed to detect significant differences between the overall effects of each moderator subgroup (Borenstein et al., 2009).

Reporting bias. In order to assess reporting bias, we completed a funnel plot analysis and applied the trim and fill method by Duval and Tweedie (2000a, 2000b) and Duval (2005), as implemented in the *metafor* package in *R* (Viechtbauer, 2010). Following the recommendations of Viechtbauer (2010), we selected the estimator "*R0*", as it provides a test of the null hypothesis that the number of missing studies on the chosen side of the funnel plot is zero. We tested this for both sides of the funnel plot.

Data management

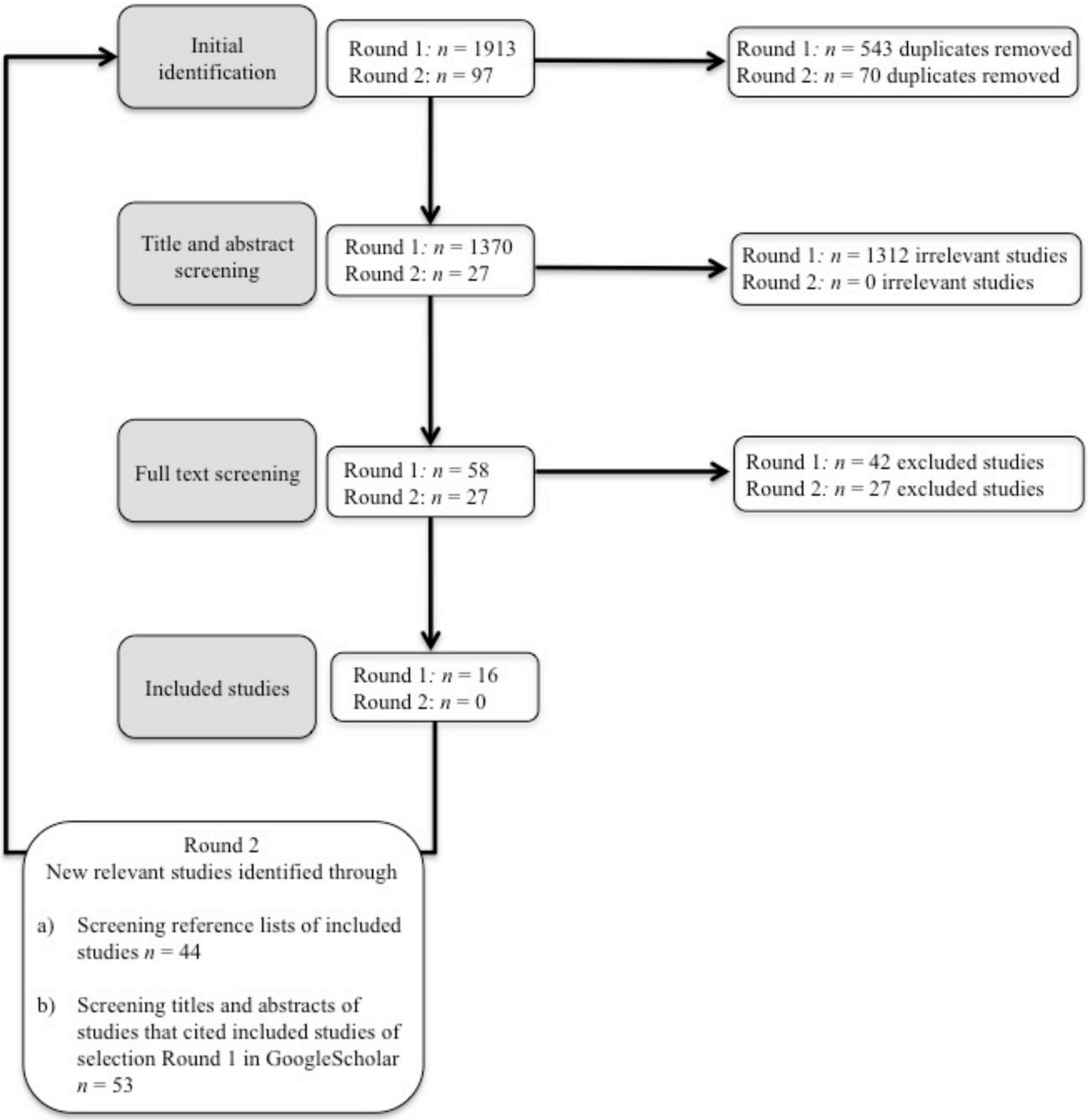
We used the software programs Mendeley and Covidence to manage our data. Mendeley served as a basis to save the search results as well as the retrieved full texts. References were imported into Covidence to complete the title and abstract as well as the full text screening. The data extraction and risk of bias assessment process were also completed in Covidence and then exported to an excel file. Meta-analyses were performed using the software program *R*.

Results

Search procedure

The search procedure is illustrated in Figure 2 and involved two rounds.

Figure 2. Flowchart of the search procedure



Note. Each step of the search process was completed by two independent reviewers and disagreements were resolved by a third reviewer.

Round 1 started with the identification of a total of 1913 study reports, of which 543 duplicates were removed. The titles and abstracts of the remaining 1370 references were screened and in total 1312 studies that did not meet inclusion criteria were excluded. Full texts of the remaining 58 studies were reviewed and 42 of them were excluded again on the basis of our inclusion criteria. Therefore, Round 1 of the search procedure ended with 16 studies meeting inclusion criteria. In *Round 2,* we additionally checked reference lists of these 16 included studies and found 44 additional references that were imported into Covidence. We also screened the titles and abstracts of studies that cited the 16 included studies in GoogleScholar and found 53 new relevant references that were also added to Covidence. Of these 97 new references, 70 duplicates were removed and for the remaining 27 studies we screened title and abstract, and the full text screening. None of these studies met inclusion criteria and thus, they did not enter the current review.

Included studies

Within the 16 studies included in this review, a total of 968 participants (404 poor readers/spellers and 564 control participants) were assessed. All studies focused on the comparison of children/adolescents with poor and typical literacy skills on at least one oral or written foreign language outcome measure. In six studies, children/adolescents with poor literacy skills were allocated to subgroups on the basis of relevant native or foreign language measures (Bekebrede et al., 2009; de Bree & Unsworth, 2014; Haisma, 2009; Helland & Kaasa, 2005; van der Leij & Morfidi, 2006; van Viersen et al., 2017). In these cases, we merged information into one data-point per study. Only Helland and Kaasa (2005) also reported the same data for the complete group of children/adolescents with poor literacy skills, so in this case we simply excluded the data-points of the subgroups.

A similar situation emerged with respect to the inclusion of more than one control group in seven studies (Bonifacci et al., 2017; Chung & Ho, 2010; de Bree & Unsworth, 2014; Ding et al., 2013; Palladino, Cismondo, Ferrari, Ballagamba & Cornoldi, 2016; van Viersen et al., 2017; Zhou et al., 2014). In two cases, in addition to age-matched control participants, reading-matched control participants were included as a separate group (Chung & Ho, 2010; Zhou et al., 2014). In order to focus on the comparison with age-matched controls, we excluded reading-matched information (McArthur et al., 2013; McDougall, Borowsky, MacKinnon & Hymel, 2005). Data on more than one control group was also excluded for Bonifacci et al. (2017) and Palladino et al. (2016), because they introduced a new comparison with a group of children with typical literacy skills with minority bilingual background and with English language difficulties, respectively. Lastly, de Bree and Unsworth (2014), Ding et al. (2013) and van Viersen et al. (2017) distinguished between good and average readers, participants from bilingual and regular schools and typically developing and gifted participants, respectively. As this distinction was not made in other studies, we merged the data of both control groups for each study.

With respect to foreign language outcome measures, two studies reported more than one data-point on the same task (van Viersen et al., 2017; Zhou et al., 2014). For van Viersen et al. (2017) we merged the data-points collected for the same English as a foreign language task at different developmental stages. However, we excluded the information on the participants' performance on German and French as a second foreign language, because all other studies only focused on foreign language outcome measures that represented the participants' first foreign language. In the case of Zhou et al. (2014), we excluded information from the first and second test point of this longitudinal study, because participants' literacy status was only determined at the third test point. Due to this moderate risk of bias in the classification to intervention status (see risk of bias assessment section for details), we only included the third data-point of this study.

Location of studies. Studies were completed in the Netherlands (six studies: Bekebrede et al., 2007; de Bree & Unsworth, 2014; Haisma, 2009; Morfidi et al., 2007; van der Leij & Morfidi, 2006; van Viersen et al., 2017), Italy (3 studies: Bonifacci et al., 2017; Palladino et al., 2013, 2016), Hong Kong (3 studies: Chung & Ho, 2010; Ho & Fong, 2005; Zhou et al., 2014), Norway (2 studies: Helland & Kaasa, 2005; Helland & Morken, 2016), Poland (1 study: Lockiewicz & Jaskulskaa, 2016) and China (1 study: Ding et al., 2013).

Participants. Relevant information on moderators related to participant characteristics is shown in Appendix 2. Only three studies reported detailed information on the native language profile of children/adolescents with poor literacy skills (Bekebrede et al., 2007; Haisma, 2007; Morfidi et al., 2007). With respect to oral versus written language deficits, Bekebrede et al. (2007) and Morfidi et al. (2007) included only children/adolescents with average oral native language skills, but poor written native language skills. None of the studies distinguished between selective reading and spelling deficits and only two studies listed information on reading difficulty subtypes. While Bekebrede et al. (2007) included children/adolescents with lexical and other reading deficits, Haisma (2007) assessed participants with sublexical and lexical reading difficulties. Information was completely missing for 13 studies.

With respect to linguistic background of participants, four studies reported that their participants had a purely monolingual background (Bonficacci et al., 2017; Chung & Ho, 2010; Ding et al., 2013; Ho & Fong, 2005). Only Morfidi et al. (2007) stated that their sample also included children with a bilingual background. We planned to allocate studies assessing pluri-/multilingual participants to the group of studies with bilingual participants. However, none of the studies matched this criteria. Information was missing for 11 studies.

Foreign language instruction. In three studies, frequency of foreign language instruction consisted of two to four classes per week (Bekebrede et al., 2007; Morfidi et al., 2007; van der Leij & Morfidi, 2006) and in two studies participants received between 30-60 minutes of instruction per class (Bekebrede et al., 2007; Morfidi et al., 2007). Information was missing for 13 studies.

Regarding language pairing, the native languages of participants were Dutch (six studies: Bekebrede et al., 2007; de Bree & Unsworth, 2014; Haisma, 2009; Morfidi et al., 2007; van der Leij & Morfidi, 2006; Van Viersen et al., 2017), Italian (3 studies: Bonficacci et al., 2017; Palladino et al., 2013, 2016), Cantonese (3 studies: Chung & Ho, 2010; Ho &

Fong, 2005; Zhou et al., 2014), Norwegian (2 studies: Helland & Kaasa, 2005, Helland & Morken, 2016), Polish (1 study: Lockiewicz & Jaskulskaa, 2016) and Mandarin (1 study: Ding et al., 2013). The foreign language assessed in all studies was English. Therefore, 12 studies focused on the combination of two Indo-European languages, with alphabetic writing systems. In all cases the native language was a predominantly regular orthography paired with a predominantly irregular foreign language orthography (Bekebrede et al., 2007; Bonficacci et al., 2017; de Bree & Unsworth, 2014; Haisma, 2009; Helland & Kaasa, 2005; Helland & Morken, 2016; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013, 2016; van der Leij & Morfidi, 2006; van Viersen et al., 2017). In contrast, in the remaining four studies the language combination was a non-Indo-European native language with an ideographic writing system with a foreign Indo-European language with an alphabetic writing system (Chung & Ho, 2010; Ding et al., 2013; Ho & Fong, 2005; Zhou et al., 2014).

Onset age of foreign language learning was early childhood (onset age before 6 years) for three studies (Chung & Ho, 2010; Ho & Fong, 2005; Zhou et al., 2014), late childhood (onset age from 6-11 years) for five studies (Ding et al., 2013; Helland & Kaasa, 2005, 2016; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007) and adolescence (onset age from 12-17 years) for three studies (Bekebrede et al., 2007; van der Leij & Morfidi, 2006; van Viersen et al., 2017). None of the studies reported the inclusion of participants with an onset age of foreign language instruction in early adulthood (onset age from 18 years onwards) and information was missing for five studies (Bonficacci et al., 2017; de Bree & Unsworth, 2014; Haisma, 2009; Palladino et al., 2013, 2016).

Foreign language assessment. The age at foreign language assessment was late childhood (6-11 years of age) for seven studies (Bonficacci et al., 2017; Chung & Ho, 2010; Ding et al., 2013; Helland & Morken, 2016; Ho & Fong, 2005; Morfidi et al., 2007; Zhou et al., 2014) and adolescence (12-17 years of age) for nine studies (Bekebrede et al., 2007; de Bree & Unsworth, 2014; Haisma, 2009; Helland & Kaasa, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2013, 2016; van der Leij & Morfidi, 2006; van Viersen et al., 2017). None of the studies assessed participants in early childhood (before 6 years of age) or early adulthood (from 18 years of age onwards).

An overview of the foreign language outcome measures collected by each study can be found in Appendix 3. No information was available on foreign language speech discrimination and production. Receptive vocabulary knowledge and word production skills were investigated by six (Bekebrede et al., 2007; de Bree & Unsworth, 2014; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006; Zhou et al., 2014) and five studies (Chung & Ho, 2010; Ding et al., 2013; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006), respectively. Only Helland and Kaasa (2005) tested sentence comprehension and production and Ho and Fong (2005) were the only authors measuring short term memory. Four studies explored phonological awareness (Chung & Ho, 2010; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006), while two studies measured letter knowledge (Chung & Ho, 2010; Morfidi et al., 2007). Word and nonword reading skills were assessed by 14 and seven studies respectively (word reading: Bekebrede et al., 2007; Bonficacci et al., 2017; Chung & Ho, 2010; de Bree & Unsworth, 2014; Ding et al., 2013; Helland & Kaasa, 2005; Helland & Morken, 2016; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006; van Viersen et al., 2017; Zhou et al., 2014; nonword reading: Bekebrede et al., 2007; Bonficacci et al., 2017; de Bree & Unsworth, 2014; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006). Seven studies gathered information on orthographic knowledge (Bekebrede et al., 2007; Chung & Ho, 2010; de Bree & Unsworth, 2014; Haisma, 2009; Morfidi et al., 2007; van der Leij & Morfidi, 2006; van Viersen et al., 2017) and four assessed reading comprehension (Bonficacci et al., 2017; Ding et al., Morfidi et al., 2007; van der Leij & Morfidi, 2006). Spelling skills were measured by eight studies (Bonficacci et al., 2017; Haisma, 2009; Helland & Kaasa, 2005; Helland &

Morken, 2016; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2016; Van Viersen et al., 2017), while only Helland and Kaasa (2005) and Helland and Morken (2016) tested translation skills.

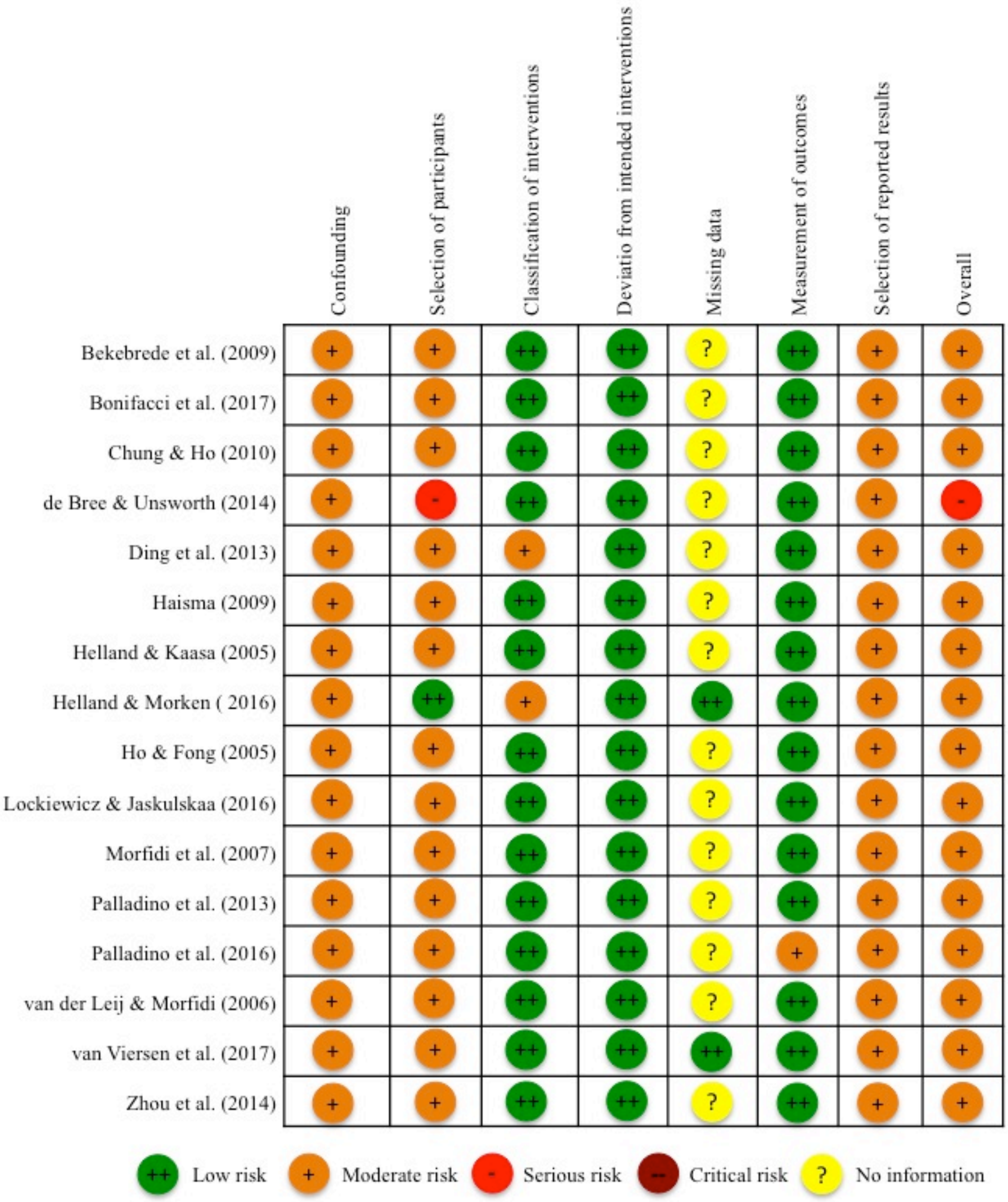
Excluded studies

We present a detailed list of studies excluded in the full text screening phase in the supplemental materials. Reasons for exclusion were (in order of frequency): access to foreign language outside of instruction context (16 studies), poor literacy skills not determined as 1 *SD*, year or grade below the expected level (5 studies), literacy performance measured by self- or teacher reports (5 studies), allocation to participant groups based on foreign language performance (4 studies), inclusion of participants in post secondary education (2 studies), no comparison with controls on any foreign language measures (2 studies), and no assessment of oral or written foreign language performance (2 studies). Furthermore, 33 studies were excluded because they represented duplicated reports on the same sample of participants (15 studies), were no empirical study reports (9 studies), focused on qualitative analyses (6 studies), were single case studies (2 studies) or access to full text was not available even after contacting authors (1 study).

Risk of bias in included studies

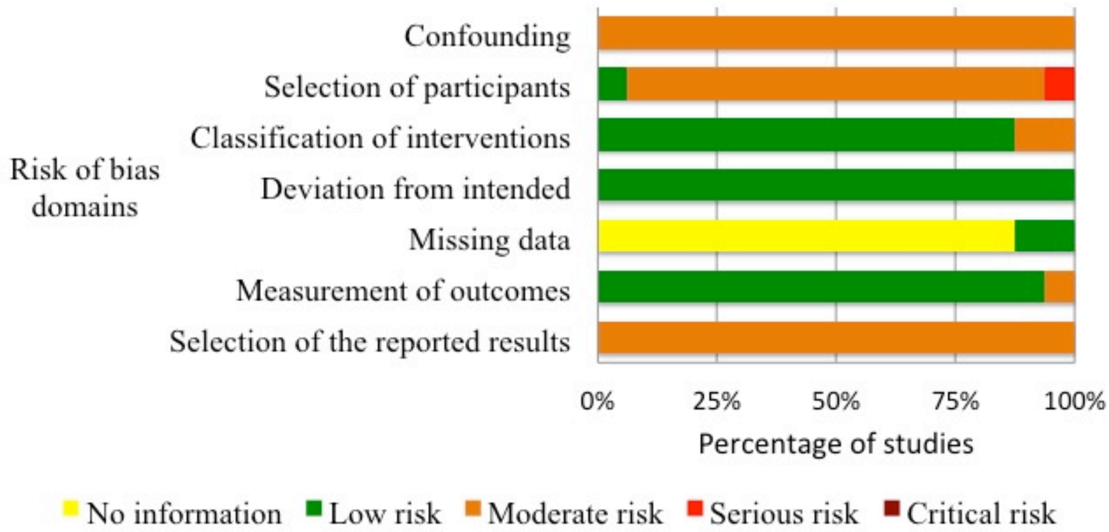
We present details on the risk of bias assessment for each study following the ROBINS I rating scale (Sterne et al., 2016) in Figure 3.

Figure 3. Risk of bias assessment of included studies



Furthermore, Figure 4 summarizes the risk of bias based on the seven domains that were assessed.

Figure 4. Overview of risk of bias of included studies (*n* = 16)



As the information contained in the study reports was not sufficient to assess all of the domains of risk of bias, in many cases we contacted authors to obtain additional information.

Confounding. As all included studies were non-randomized control studies, by default confounding was expected, because different baseline characteristics of participants could have influenced the results (Sterne et al., 2016). Therefore, all studies had at least a moderate risk of bias in this domain (Sterne et al., 2016). In order to maintain a moderate risk of bias judgement, studies had to report information on the measurement and control of important confounding domains such as age, socio-economic status (SES), nonverbal reasoning, oral native language skills, etc. Furthermore, the reliability and validity of the measurement of these variables had to be sufficient to not expect serious confounding (Sterne et al., 2016). All studies fulfilled these conditions; either through pairwise matching between experimental and control participants or through statistical adjustment with Bayesian analysis or ANCOVAs. Therefore, all studies received a moderate risk of bias judgement with respect to confounding.

Selection of participants into the study. Two conditions had to be fulfilled to reflect a low risk of bias in this domain. First, the start of foreign language instruction and also the timing of foreign language assessment had to be the same for most participants (Sterne et al., 2016). All included studies fulfilled this condition. Second, the selection of participants had to be unrelated to their foreign language attainment (Sterne et al., 2016). This condition was only fulfilled by Helland and Morken (2016), as they explicitly stated that parents of all children in the participating schools were contacted. Therefore, this study received a low risk of bias judgement. In contrast, all of the other studies either did not report how they selected participants into the study (Bonficacci et al., 2017; Chung & Ho, 2010; Ding et al., 2013; Haisma, 2009; Helland & Kaasa, 2005; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2013, 2016; Zhou et al., 2014) or mentioned that school staff (e.g. counsellors, teachers, etc.) had selected the participants (Bekebrede et al., 2007; de Bree & Unsworth, 2014; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2013, 2016; van Viersen et al., 2017; Zhou et al., 2014). This indicates a moderate risk of bias, because school staff could have selected participants based on their knowledge of the participants' foreign language performance. However, all of the studies, except de Bree and Unsworth (2014), confirmed the literacy status of each of the participants either through an external diagnosis of poor literacy skills or by administering a literacy test within the context of the study. In contrast, de Bree and Unsworth (2014) solely relied on the information of school staff for the selection of participants into the study. Although the authors acknowledged the presence of selection bias in their study in a footnote, this still represents a serious risk of bias. Therefore, we excluded this study from further analysis.

Classification of interventions. Two aspects were crucial in this risk of bias domain. First, intervention status had to be well defined (Sterne et al., 2016). Children/adolescents with poor literacy skills had to show a performance of at least 1 *SD*, year or grade below the expected level on one or more of the following measures: word/nonword reading accuracy, reading fluency and/or spelling. While all studies fulfilled this condition, Ding et al. (2013) and Helland and Morken (2016) showed a moderate risk of bias. The group of children/adolescents with poor literacy skills in Ding et al.´s (2013) study was selected because they scored 1 *SD* below the mean of the study sample itself (n = 102). This represents a moderate risk, because the identification of children/adolescents with poor literacy skills could have been biased by the characteristics of the sample of total participants from which they were selected. Helland and Morken (2016) identified poor literacy skills based on a below average performance (< 25th percentile) on at least two out of four literacy measures. While for three of these measures independent standardized test norms were available, one (i.e. text reading) was developed by the authors (see Helland, Plante & Hugdahl, 2011, for additional information to Helland & Morken (2016)). For this measure the cut-off criterion (< 25th percentile) was based on the sample of the study (n = 42). Therefore, similar to Ding et al. (2013), the identification of children/adolescents with poor literacy skills could have been biased by the characteristics of the specific study sample. Based on this information, we assigned a moderate risk of bias to Helland and Morken (2016).

A second crucial aspect with respect to the classification of interventions was that the assignment of intervention status should not have been determined retrospectively (Sterne et al., 2016). Although two studies were especially vulnerable with respect to this condition, after a careful analysis, we decided that no risk of bias was present. First, Zhou et al. (2014) completed a longitudinal study in which poor literacy status was defined at the last testing point and retrospectively assigned to two previous testing points at earlier developmental stages. With this risk of bias domain in mind, we had already excluded the first two previous testing points during data extraction. Therefore, no risk of bias was present for the third data-point which was included in this review. Similarly, in their longitudinal study, Helland and Morken (2016) determined poor literacy status at the last testing point and retrospectively assigned to two previous testing points at earlier developmental stages. However, the authors only reported on foreign language outcome measures for the third testing point, when poor literacy status was defined. Therefore, no risk of bias was identified.

Deviation from intended interventions. All studies showed a low risk of bias in this domain. Lockiewicz and Jaskulskaa (2016) were the only authors who reported information on a potential risk of bias of deviation from intended interventions in the form of extra-curricular private foreign language tutoring. Furthermore, foreign language teachers could have provided educational accommodations to children/adolescents with poor literacy skills if they had knowledge on their students' native language performance (e.g. dyslexia diagnosis). However, both situations of potential risk of bias reflect usual practice and are therefore assigned a low risk of bias (Sterne et al., 2016).

Missing data. We were not able to assess this risk of bias domain for most studies, because no explicit information was given in the studies, with the exception of Helland and Morken (2016) and van Viersen et al. (2017). A low risk was assigned to studies where the number of participants in the results matched the number of participants in the methods section. However, other studies did not provide the number of participants when presenting results.

Measurement of outcomes. All studies showed a low risk of bias with the exception of Palladino et al. (2016). In this study spelling errors were scored according to a pre-defined grid. Since no information was available regarding the reliability of this measure, we assigned a moderate risk of bias.

Selection of reported results. We identified a moderate risk of bias for all studies because no pre-registered protocols or statistical analysis plans were available for any of the studies (Sterne et al., 2016). However, the information on outcome measurements in the methods and results section of each study report was consistent.

Overall risk of bias. We determined the overall risk of bias as the highest risk of bias judgement received by a study in any of the domains (Sterne et al., 2016). This was a moderate risk of bias for all studies, with the exception of de Bree and Unsworth (2014). In this study, we found a serious risk of bias in the selection of participants and therefore assigned an overall serious risk of bias. This led to the exclusion of this study from further analyses.

Meta-analyses of foreign language outcome measures

Separate random effects modelling meta-analyses for 10 out of 15 foreign language outcome measures could be computed. Analyses for the remaining five foreign language outcome measures were not possible due to limited data (less than 2 study reports). Results are presented in Table 3.

Table 3

Meta-analytic results

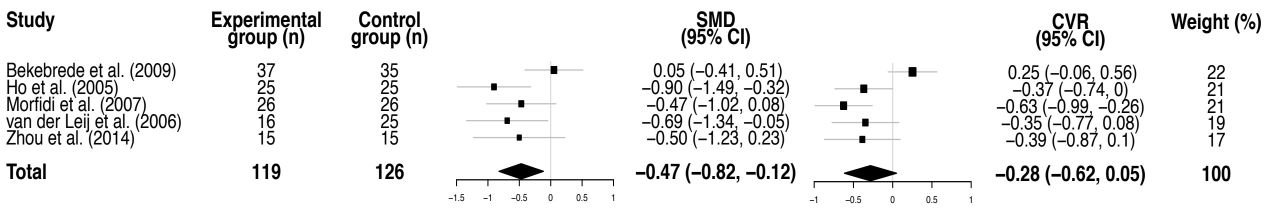
FL outcome measure	Number of studies	Participant groups: Poor readers/ spellers (n)	Participant groups: Control participants (n)	Mean difference between groups: Q	I^2	*SMD*	95% CI	Z	Variance difference between groups: Q	I^2	*CVR*	95% CI	Z
Receptive vocabulary knowledge	5	119	126	7.38^{ns}	45.81%	-0.47	[-0.82, -0.12]	-2.59^{ns}	15.01*	73.35%	-0.28	[-0.62, 0.05]	-1.66^{ns}
Spoken word production	4	104	104	7.80^{ns}	61.56%	-1.10	[-1.60, -0.68]	-4.85*	11.01*	72.76%	-0.44	[-0.79, -0.09]	-2.47^{ns}
Phonological awareness	4	95	104	3.87^{ns}	22.47%	-1.10	[-1.40, -0.76]	-6.65*	2.64^{ns}	13.54%	-0.38	[-0.56, -0.20]	-4.12*
Letter knowledge	2	54	54	2.61^{ns}	61.67%	-1.23	[-1.90, -0.56]	-3.59*	40.68*	97.54%	-0.81	[-2.40, 0.83]	-0.96^{ns}
Word reading	13	319	455	114.33*	89.50%	-1.60	[-2.10, -1.10]	-6.47*	60.43*	80.14%	-0.56	[-0.78, -0.35]	-5.08*
Nonword reading	6	169	235	18.46*	72.92%	-0.98	[-1.37, -0.58]	-4.85*	11.20*	55.37%	-0.22	[-0.42, -0.03]	-2.22^{ns}
Orthographic knowledge	6	160	155	30.18*	83.43%	-1.40	[-1.99, -0.84]	-4.82*	77.96*	93.58%	-0.70	[-1.25, -0.15]	-2.47^{ns}
Reading comprehension	4	79	211	2.97^{ns}	0%	-1.00	[-1.29, -0.75]	-7.35*	5.83^{ns}	48.58%	-0.23	[-0.48, 0.02]	-1.82^{ns}
Spelling	8	211	308	25.54*	72.59%	-1.40	[-1.77, -1.00]	-7.51*	16.31*	57.09%	-0.38	[-0.57, -0.19]	-3.99*
Translation	2	33	48	4.25*	76.47%	-1.20	[-2.28, -0.20]	-2.34^{ns}	2.62^{ns}	61.92%	-0.79	[-1.30, -0.29]	-3.07*

Note. FL = Foreign language; SMD = Standardized Mean Difference; CVR = natural logarithm of the ratio of variation coefficients (Nakagawa et al., 2015). The significance level was $p < .05$ for the Q statistic and $p < .005$ for the Z statistic (Bonferroni correction for 10 comparisons). Cells marked in grey show overall effects that could not be interpreted due to the presence of significant between study heterogeneity

Meta-analytic results revealed overall *SMDs* and *CVRs* for 10 foreign language outcome measures. However, only five out of 10 overall *SMDs* could be interpreted, due to significant heterogeneity between study effects (*Q* statistic $p < .05$). The interpretable effects concerned foreign language receptive vocabulary knowledge, spoken word production, phonological awareness, letter knowledge and reading comprehension. For foreign language receptive vocabulary knowledge, children/adolescents with poor literacy skills on average showed a similar performance to their peers with typical literacy skills. In contrast, a lower attainment of poor readers/spellers was found for foreign language spoken word production, phonological awareness, letter knowledge and reading comprehension. However, since results based on *SMD*s are derived from group averages they may not take into consideration individual differences within each participant group. Therefore, in addition to overall *SMDs*, we computed overall *CVRs*. This complementary analysis allowed us to estimate to what extent results from overall *SMD*s were likely to vary across individual poor readers/spellers. Due to significant heterogeneity between study effects (*Q* statistic $p < .05$), *CVR*s could only be interpreted with respect to two foreign language outcome measures, namely for phonological awareness and reading comprehension. According to these results poor readers/spellers varied significantly more than control participants in their foreign language phonological awareness performance. In contrast, performance on foreign language reading comprehension varied to a similar extent across participant groups. Below, we detail the results for each outcome measure separately.

Receptive vocabulary knowledge. No overall significant difference between the average foreign language receptive vocabulary knowledge of children/adolescents with poor and typical literacy skills was found, $SMD = -\ 0.47$, 95% *CI* (-0.82, -0.12); $Z = -2.59$, $p = .009^{ns}$. Heterogeneity between *SMDs* was non-significant and low (Higgins et al., 2003), $Q = 7.38$, $df = 4$, $p = 0.12^{ns}$; $\tau^2 = 0.07$; $I^2 = 45.81\%$, 95% *CI* (0.00%, 80.14%). The overall variation between foreign language receptive vocabulary knowledge in both participant groups could not be interpreted, as *CVR*s were significantly heterogeneous, *CVR* = - 0.28, 95% *CI* (-0.62, -0.05); Z = -1.66, p = .096[ns]; Q = 15.01, df = 4, p = .004; τ^2 = 0.10; I^2 = 73.35%, 95% *CI* (33.49%, 89.32%). Analyses were based on five studies (Bekebrede et al., 2009; Ho & Fong, 2005; Morfidi et al., 2007, van der Leij & Morfidi, 2006, Zhou et al., 2014). Results are depicted in Figure 5.

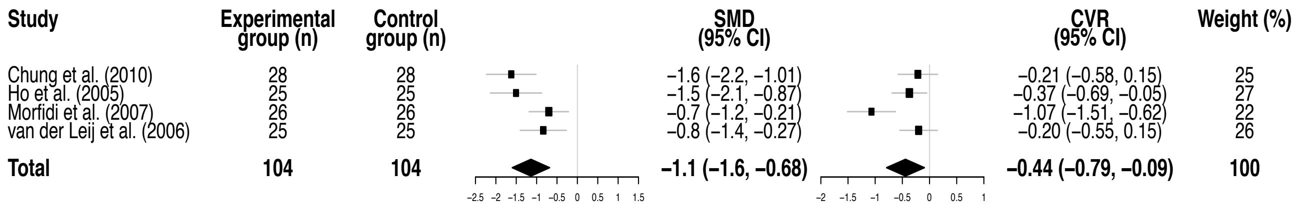
Figure 5. Meta-analysis on foreign language receptive vocabulary knowledge



Note. Heterogeneity *SMD*: Q = 7.38, df = 4, p = .12; τ^2 = 0.07; I^2 = 45.81%, 95% *CI* (0.00%, 80.14%); Overall effect: Z = -2.59, p = .009; Heterogeneity *CVR*: Q = 15.01, df = 4, p = .004; τ^2 = 0.10; I^2 = 73.35%, 95% *CI* (33.49%, 89.32%); Overall effect: Z = -1.66, p = .096. The significance level was $p < .05$ for the Q statistic and p < .005 for the Z statistic (Bonferroni correction for 10 comparisons).

Spoken word production. We found a significant overall difference of *SMD* = -1.1, 95% *CI* (-1.60, -0.68); Z = -4.85, $p < .001$ between children/adolescents with poor and typical literacy skills. Non-significant moderate heterogeneity was observed, Q = 7.80, df = 3, p = .05; τ^2 = 0.13; I^2 = 61.56%, 95% *CI* (0.00%, 87.11%). Furthermore, overall differences in the variation of the performance of both participant groups could not be interpreted, as *CVRs* significantly varied across studies, *CVR* = - 0.44, 95% *CI* (-0.79, -0.09); Z = -2.47, p = .013[ns]); Q = 11.01, df = 3, p = .011; τ^2 = 0.09; I^2 = 72.76%, 95% *CI* (23.15%, 90.34%). These results are based on the analyses of four studies (Bekebrede et al., 2009; Ho & Fong, 2005; Morfidi et al., 2007, van der Leij & Morfidi, 2006, Zhou et al., 2014) (see Figure 6).

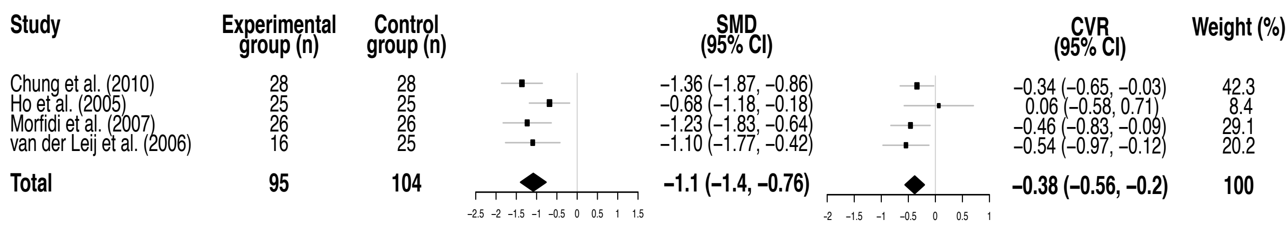
Figure 6. Meta-analysis on foreign language spoken word production



Note. Heterogeneity *SMD*: $Q = 7.80$, $df = 3$, $p = .05$; $\tau^2 = 0.13$; $I^2 = 61.56\%$, 95% *CI* (0.00%, 87.11%); Overall effect: $Z = -4.85$, $p < .001$. Heterogeneity *CVR*: $Q = 11.01$, $df = 3$, $p = .01$; $\tau^2 = 0.09$; $I^2 = 72.76\%$, 95% *CI* (23.15%, 90.34%); Overall effect: $Z = -2.47$, $p = .01$. The significance level was $p < .05$ for the Q statistic and $p < .005$ for the Z statistic (Bonferroni correction for 10 comparisons).

Phonological awareness. The overall difference between groups was $SMD = -1.1$, 95% *CI* (-1.40, -0.76); $Z = -6.65$, $p < .001$. Results revealed non-significant low heterogeneity between study effects, $Q = 3.87$, $df = 3$, $p = .28$; $\tau^2 = 0.02$; $I^2 = 22.47\%$, 95% *CI* (0.00%, 88.13%). However, a significant overall difference between the variation of performance in each participant group showed that poor readers/spellers varied significantly more than control participants on foreign language phonological awareness measures, $CVR = -0.38$, 95% *CI* (-0.56, -0.20); $Z = -4.12$, $p < .001$; $Q = 2.64$, $df = 3$, $p = .45^{ns}$; $\tau^2 = 0.01$; $I^2 = 13.54\%$, 95% *CI* (0%, 82.61%). Analyses were based on four studies (Chung & Ho, 2010; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006) (see Figure 7).

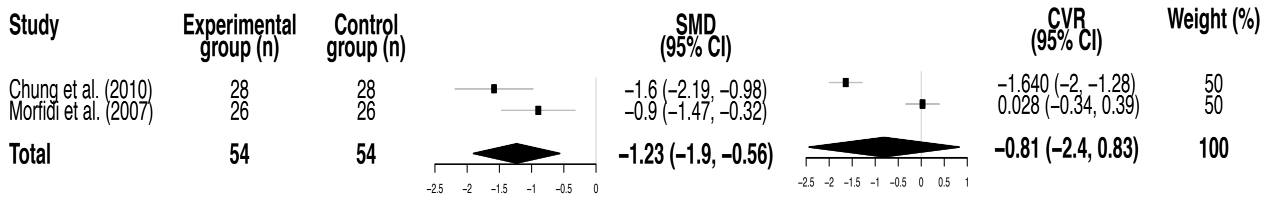
Figure 7. Meta-analysis on foreign language phonological awareness



Note. Heterogeneity *SMD*: $Q = 3.87$, $df = 3$, $p = .28$; $\tau^2 = 0.02$; $I^2 = 22.47\%$, 95% *CI* (0.00%, 88.13%); Overall effect: $Z = -6.65$, $p < .001$. Heterogeneity *CVR*: $Q = 2.64$, $df = 3$, $p = .45$; $\tau^2 = 0.01$; $I^2 = 13.54\%$, 95% *CI* (0.00%, 82.61%); Overall effect: $Z = -4.12$, $p < .001$. The significance level was $p < .05$ for the Q statistic and $p < .005$ for the Z statistic (Bonferroni correction for 10 comparisons).

Letter knowledge. Children/adolescents with poor literacy skills on average performed significantly lower on letter knowledge tasks than their peers with typical literacy skills with a *SMD* = -1.23, 95% *CI* (-1.90, -0.56); $Z = -3.59$, $p < .001$. Moderate heterogeneity between study effects did not reach statistical significance, $Q = 2.61$, $df = 1$, $p = 0.11^{ns}$; $\tau^2 = 0.14$; $I^2 = 61.67\%$, 95% *CI* (0.00%, 91.13%). The overall *CVR* could not be interpreted due to significant heterogeneity across study effects, *CVR* = - 0.81, 95% *CI* (-2.40, -0.83; $Z = -0.96$, $p = .33^{ns}$; $Q = 40.68$, $df = 1$, $p < .001$; $\tau^2 = 1.35$; $I^2 = 97.54\%$, 95% *CI* (93.92%, 99.00%). These results are based on two studies (Chung & Ho, 2010; Morfidi et al., 2007) (see Figure 8).

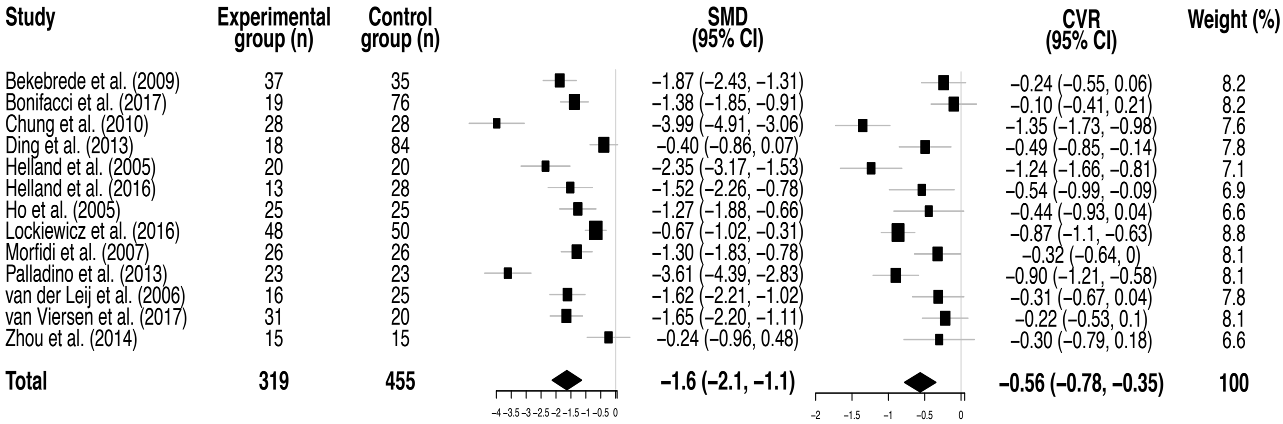
Figure 8. Meta-analysis on foreign language letter knowledge



Note. Heterogeneity *SMD*: $Q = 2.61$, $df = 1$, $p = .11$; $\tau^2 = 0.14$; $I^2 = 61.67\%$, 95% *CI* (0.00%, 91.13%); Overall effect: $Z = -3.59$, $p < .001$. Heterogeneity *CVR*: $Q = 40.68$, $df = 1$, $p < .001$; $\tau^2 = 1.35$; $I^2 = 97.54\%$, 95% *CI* (93.92%, 99.00%); Overall effect: $Z = -0.96$, $p = .33$. The significance level was $p < .05$ for the Q statistic and $p < .005$ for the Z statistic (Bonferroni correction for 10 comparisons).

Word Reading. The overall *SMD* was not interpretable, due to significantly high levels of heterogeneity between study effects, *SMD* = -1.60, 95% *CI* (-2.10, -1.10); $Z = -6.47$, $p < .001$; $Q = 114.33$, $df = 12$, $p < .001$; $\tau^2 = 0.72$; $I^2 = 89.50\%$, 95% *CI* (83.90%, 93.16%). Similarly, the overall *CVR* was not interpreted due to significant between-study heterogeneity, *CVR* = - 0.56, 95% *CI* (-0.78, -0.35); $Z = -5.08$, $p < .001$; $Q = 60.43$, $df = 12$; $p < .001$; $\tau^2 = 0.12$; $I^2 = 80.14\%$, 95% *CI* (66.89%, 88.09%). As this was the only foreign language outcome measure for which data from more than 10 study records was available, we further investigated the heterogeneity previously mentioned through moderator analysis (see section moderator analysis). Results were based on thirteen studies (Bekebrede et al., 2007; Bonficacci et al., 2017; Chung & Ho, 2010; Ding et al., 2013; Helland & Kaasa, 2005; Helland & Morken, 2016; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006; van Viersen et al., 2017; Zhou et al., 2014) (see Figure 9).

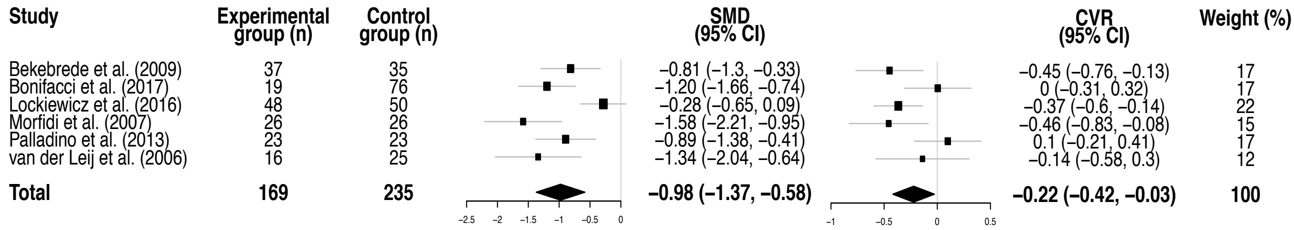
Figure 9. Meta-analyses on foreign language word reading



Note. Heterogeneity *SMD*: $Q = 114.33$, $df = 12$, $p < .001$; $\tau^2 = 0.72$; $I^2 = 89.50\%$, 95% *CI* (83.90%, 93.16%); Overall effect: $Z = -6.47$, $p < .001$. Heterogeneity *CVR*: $Q = 60.43$, $df = 12$, $p < .001$; $\tau^2 = 0.12$; $I^2 = 80.14\%$, 95% *CI* (66.89%, 88.09%); Overall effect: $Z = -5.08$, $p < .001$. The significance level was $p < .05$ for the Q statistic and $p < .005$ for the Z statistic (Bonferroni correction for 10 comparisons).

Nonword reading. Neither the overall *SMD* nor the *CVR* for foreign language nonword reading could be interpreted due to significantly high heterogeneity: $SMD = -0.98$, 95% *CI* (-1.37, -0.58); $Z = -4.85$, $p < .001$; $Q = 18.46$, $df = 5$, $p = .002$; $\tau^2 = 0.17$; $I^2 = 72.92\%$, 95% *CI* (37.76%, 88.22%); $CVR = -0.22$, 95% *CI* (-0.42, -0.03); $Z = -2.22$, $p = .02^{ns}$; $Q = 11.20$, $df = 5$, $p = .04$; $\tau^2 = 0.03$; $I^2 = 55.37\%$, 95% *CI* (0%, 82.07%). Analyses were based on six studies (Bekebrede et al., 2007; Bonficacci et al., 2017; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006) (see Figure 10).

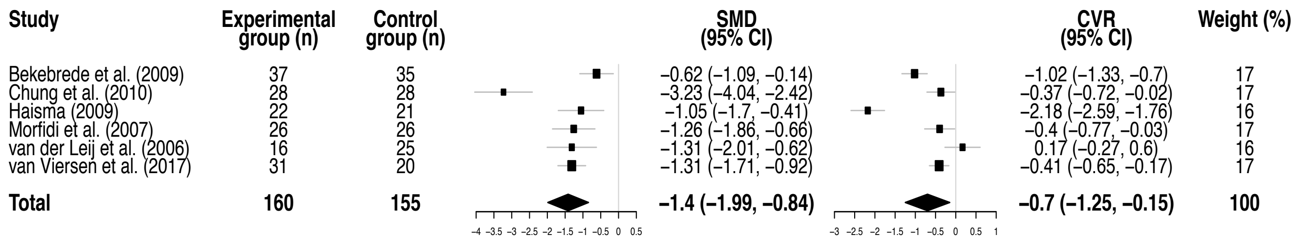
Figure 10. Meta-analysis on foreign language nonword reading



Note. Heterogeneity *SMD*: $Q = 18.46$, $df = 5$, $p = .002$; $\tau^2 = 0.17$; $I^2 = 72.92\%$, 95% *CI* (37.76%, 88.22%); Overall effect: $Z = -4.85$, $p < .001$. Heterogeneity *CVR*: $Q = 11.20$, $df = 5$, $p = .04$; $\tau^2 = 0.03$; $I^2 = 55.37\%$, 95% *CI* (0.00%, 82.07%); Overall effect: $Z = -2.22$, $p = .02$. The significance level was $p < .05$ for the *Q* statistic and p < .005 for the *Z* statistic (Bonferroni correction for 10 comparisons).

Orthographic knowledge. Once again, overall *SMD* and *CVR* could not be interpreted due to the presence of significant heterogeneity, $SMD = -1.40$ (95% *CI* -1.99, -0.84), $Z = -4.82$, $p < .001$; $Q = 30.18$, $df = 5$, $p < .001$; $\tau^2 = 0.42$; $I^2 = 83.43\%$, 95% *CI* (65.32%, 92.09%); $CVR = -0.70$ (95% *CI* -1.25, -0.15), $Z = -2.47$, $p = .01^{ns}$; $Q = 77.96$, $df = 5$, $p < .001$; $\tau^2 = 0.44$; $I^2 = 93.58\%$, 95% *CI* (88.68%, 96.36%). Six studies were taken into account to obtain these results (Bekebrede et al., 2007; Chung & Ho, 2010; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006; van Viersen et al., 2017) (see Figure 11).

Figure 11. Meta-analysis on foreign language orthographic knowledge

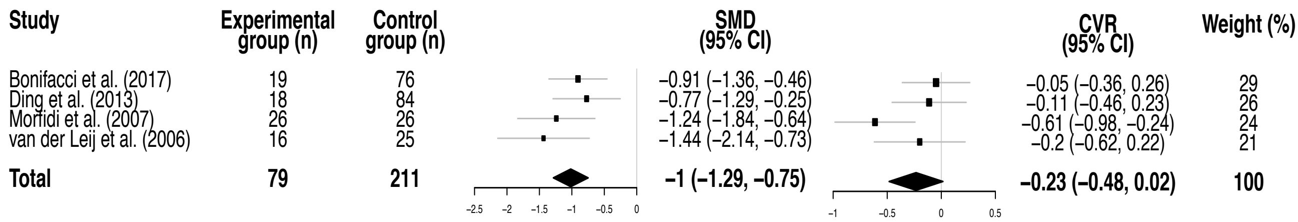


Note. Heterogeneity *SMD*: $Q = 30.18$, $df = 5$, $p < .001$; $\tau^2 = 0.42$; $I^2 = 83.43\%$, 95% *CI* (65.32%, 92.09%); Overall effect: $Z = -4.82$, $p < .001$. Heterogeneity *CVR*: $Q = 77.96$, $df = 5$, $p < .001$; $\tau^2 = 0.44$; $I^2 = 93.58\%$, 95% *CI* (88.68%, 96.36%); Overall effect: $Z = -2.47$, $p = .01$. The significance level was $p < .05$ for the *Q* statistic and p < .005 for the *Z* statistic (Bonferroni correction for 10 comparisons).

Reading comprehension. Children/adolescents with poor literacy skills on average were significantly poorer on foreign language reading comprehension tasks as compared to their peers with typical literacy skills, $SMD = -1.00$, 95% *CI* (-1.29, -0.75); $Z = -7.35$, $p <$

.001. Non-significant low levels of heterogeneity between *SMDs* were found, $Q = 2.97$, $df = 3$, $p = .39$; $\tau^2 = 0.00$; $I^2 = 0.00\%$, 95% *CI* (00.00%, 84.52%). Furthermore, a non-significant overall *CVR* in the absence of significant between-study heterogeneity indicated that the variation in performance was similar across participant groups, $CVR = -0.23$, 95% *CI* (-0.48, 0.02); $Z = -1.82$, $p = .06^{ns}$; $Q = 5.83$, $df = 3$, $p = .11$; $\tau^2 = 0.03$; $I^2 = 48.58\%$, 95% *CI* (0%, 82.96%) Analyses are based on four studies (Bonifacci et al., 2017, Ding et al., 2013, Morfidi et al., 2007; van der Leij & Morfidi, 2006) (see Figure 12).

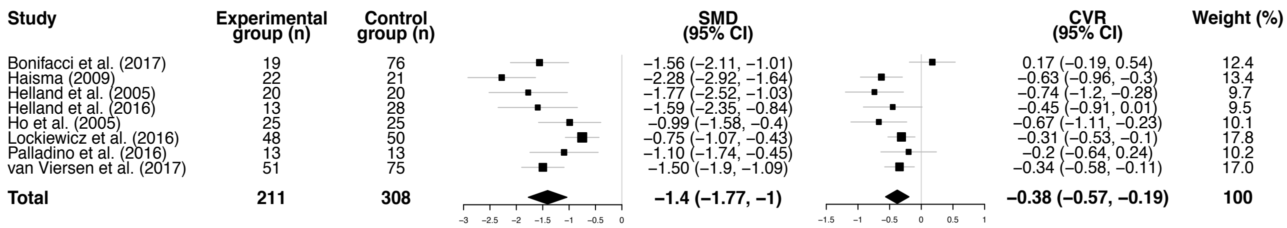
Figure 12. Meta-analysis on foreign language reading comprehension



Note. Heterogeneity *SMD*: $Q = 2.97$, $df = 3$, $p = .39$; $\tau^2 = 0.00$; $I^2 = 0\%$, 95% *CI* (0.00%, 84.52%); Overall effect: $Z = -7.35$, $p < .001$. Heterogeneity *CVR*: $Q = 5.83$, $df = 3$, $p = .11$; $\tau^2 = 0.03$; $I^2 = 48.58\%$, 95% *CI* (0.00%, 82.96%); Overall effect: $Z = -1.82$, $p = .06$. The significance level was $p < .05$ for the *Q* statistic and p < .005 for the *Z* statistic (Bonferroni correction for 10 comparisons).

Spelling. We were unable to interpret overall effects for foreign language spelling due to the presence of significant heterogeneity across study effects: $SMD = -1.40$, 95% *CI* (-1.77, -1.00), $Z = -7.51$, $p < .001$; $Q = 25.54$, $df = 7$, $p < .001$; $\tau^2 = 0.19$; $I^2 = 72.59\%$, 95% *CI* (43.90%, 86.61%); $CVR = -0.38$, 95% *CI* (-0.57, -0.19), $Z = -3.99$, $p < .001$; $Q = 16.31$, $df = 7$, $p = .02$; $\tau^2 = 0.03$; $I^2 = 57.09\%$, 95% *CI* (5.72%, 80.47%). These results are based on eight studies (Bonficacci et al., 2017; Haisma, 2009; Helland & Kaasa, 2005; Helland & Morken, 2016; Ho & Fong, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2016; van Viersen et al., 2017) (see Figure 13).

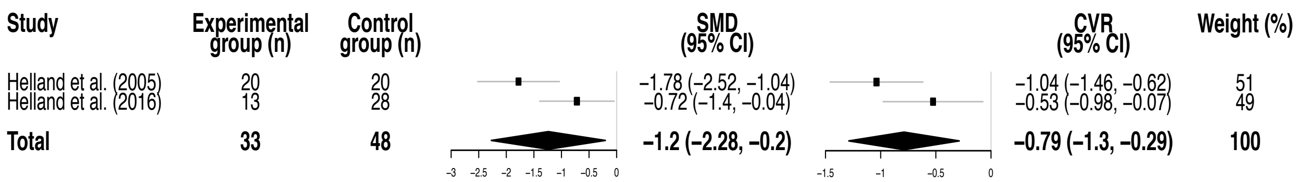
Figure 13. Meta-analysis on foreign language spelling



Note. Heterogeneity *SMD*: $Q = 25.54$, $df = 7$, $p < .001$; $\tau^2 = 0.19$; $I^2 = 72.59\%$, 95% *CI* (43.90%, 86.61%); Overall effect: $Z = -7.51$, $p < .001$. Heterogeneity *CVR*: $Q = 16.31$, $df = 7$, $p = .02$; $\tau^2 = 0.03$; $I^2 = 57.09\%$, 95% *CI* (5.72%, 80.47%); Overall effect: $Z = -3.99$, $p < .001$. The significance level was $p < .05$ for the *Q* statistic and p < .005 for the *Z* statistic (Bonferroni correction for 10 comparisons).

Translation. The overall *SMD* between groups could not be interpreted, as significantly high levels of heterogeneity were identified, $SMD = -1.20$, 95% *CI* (-2.28, -0.20), $Z = -2.34$, $p = .02^{ns}$; $Q = 4.25$, $df = 1$, $p = .04$; $\tau^2 = 0.43$; $I^2 = 76.47\%$, 95% *CI* (0%, 94.65%). Nevertheless, a significant overall *CVR* in the absence of significant heterogeneity indicated that the performance of poor readers/spellers varied to a greater extent than the performance of control participants, $CVR = -0.79$, 95% *CI* (-1.30, -0.29), $Z = -3.07$, $p = .002$; $Q = 2.62$, $df = 1$, $p = .10^{ns}$; $\tau^2 = 0.08$; $I^2 = 61.92\%$, 95% *CI* (0%, 91.20%). Results are based on two studies (Helland & Kaasa, 2005; Helland & Morken, 2016) (see Figure 14).

Figure 14. Meta-analysis on foreign language translation



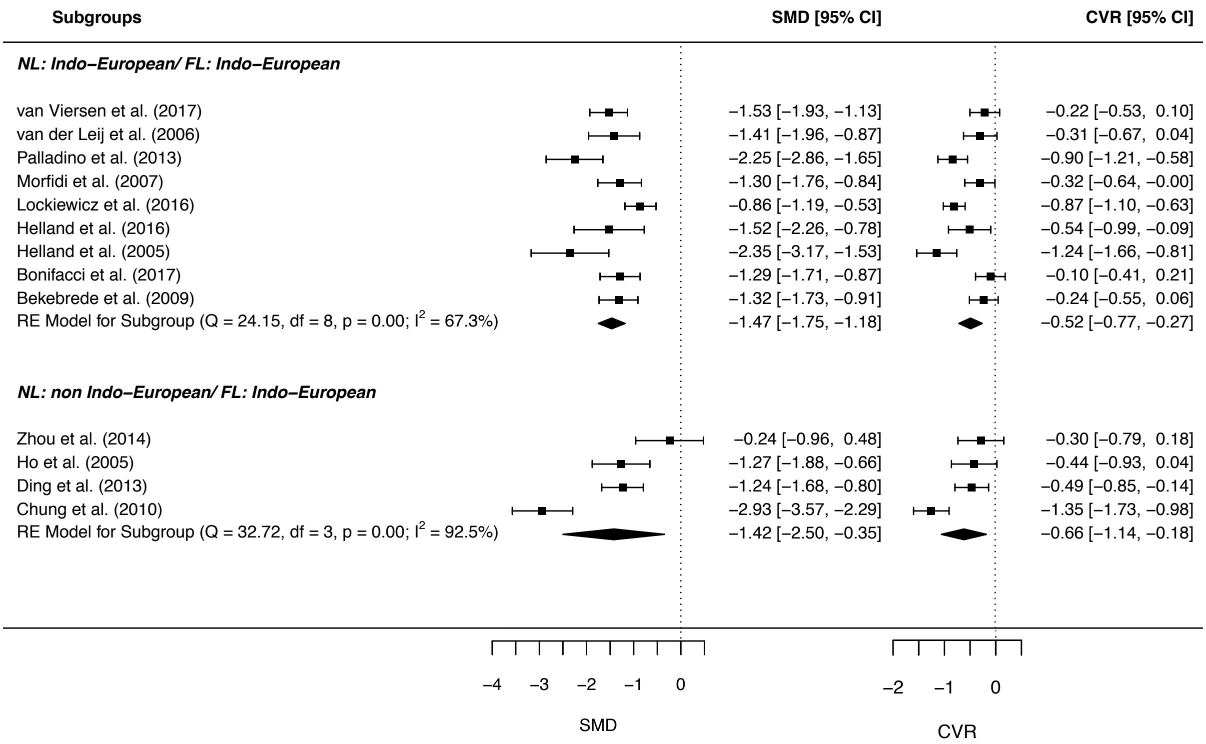
Note. Heterogeneity *SMD*: $Q = 4.25$, $df = 1$, $p = .04$; $\tau^2 = 0.43$; $I^2 = 76.47\%$ 95% *CI* (0.00%, 94.65%); Overall effect: $Z = -2.34$, $p = .02$). Heterogeneity *CVR*: $Q = 2.62$, $df = 1$, $p = .10$; $\tau^2 = 0.08$; $I^2 = 61.92\%$ 95% *CI* (0.00%, 91.20%); Overall effect: $Z = -3.07$, $p = .002$. The significance level was $p < .05$ for the *Q* statistic and p < .005 for the *Z* statistic (Bonferroni correction for 10 comparisons).

Moderator analyses

Due to serious data limitations (less than 10 studies), only three out of 11 planned moderator analyses (see Table 2) could be conducted on only one of the 15 foreign language outcome measures planned for this review (Littell et al., 2008). Thus, we focused on investigating whether (a) language pairing between native and foreign language, (b) onset age of foreign language instruction and (c) age at foreign language assessment provided a better understanding of the heterogeneity observed between study effects in foreign language word reading. An adjusted significance level of $p < .003$ was applied to correct for the five additional comparisons (on top of the 10 previous comparisons) that were computed as part of the moderator analyses (Borenstein et al., 2009).

Language pairing between native and foreign language. We compared the results of separate random effects modelling meta-analyses for studies with two types of language pairings: (a) a native Indo-European language with an alphabetic writing system combined with a foreign language with the same characteristics (Bekebrede et al., 2007; Bonifacci et al., 2017; Helland & Kaasa, 2005; Helland & Morken, 2016; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006; van Viersen et al., 2017), and (b) a native non Indo-European language with an ideographic writing system combined with a Indo-European foreign language with an alphabetic writing system (Chung & Ho, 2010; Ding et al., 2013; Ho & Fong, 2005; Zhou et al., 2014). No significant impact of this moderator was detected neither for the *SMD*, nor for the *CVR* of the foreign language word reading performance between children/adolescents with poor and typical literacy skills (*SMD*: $Z_{Diff^*} = 0.08$, $p = .94^{ns}$; *CVR*: $Z_{Diff^*} = 0.50$, $p = .61^{ns}$). Figure 15 shows the results.

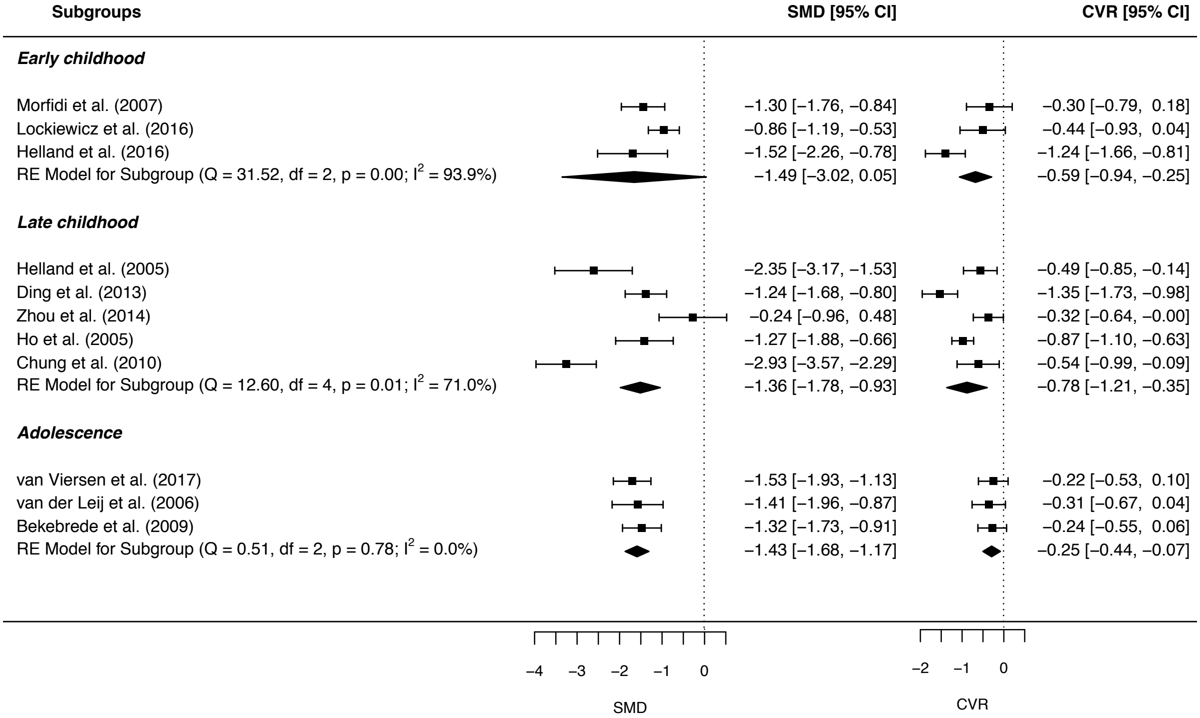
Figure 15. Moderator analysis on language pairing for foreign language word reading



Note. SMD = Standardized Mean Difference*; CVR* = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

Onset age of foreign language instruction. Subgroups of studies assessing participants with an onset age of foreign language instruction in early childhood (before age 6 - Chung & Ho, 2010; Ho & Fong, 2005; Zhou et al., 2014), late childhood (age 6 to 11 - Ding et al., 2013; Helland & Kaasa, 2005; Helland & Morken, 2016; Lockiewicz & Jaskulskaa, 2016; Morfidi et al., 2007) and adolescence (age 12 to 17 - Bekebrede et al., 2007; van der Leij & Morfidi, 2006; van Viersen et al., 2017) were compared. Results revealed no significant difference between subgroups for *SMD*, or for *CVR* (early childhood vs. late childhood: *SMD*: $Z_{Diff*} = 0.17$, $p = .86^{ns}$; *CVR*: $Z_{Diff*} = 0.63$, $p = .52^{ns}$; early childhood vs. adolescence: *SMD*: $Z_{Diff*} = 0.04$, $p = .97^{ns}$; *CVR*: $Z_{Diff*} = 2.22$, $p = .02^{ns}$; late childhood vs. adolescence: *SMD*: $Z_{Diff*} = 0.52$, $p = .60^{ns}$; *CVR*: $Z_{Diff*} = 2.70$, $p = .006^{ns}$). Figure 16 shows the corresponding forest plots.

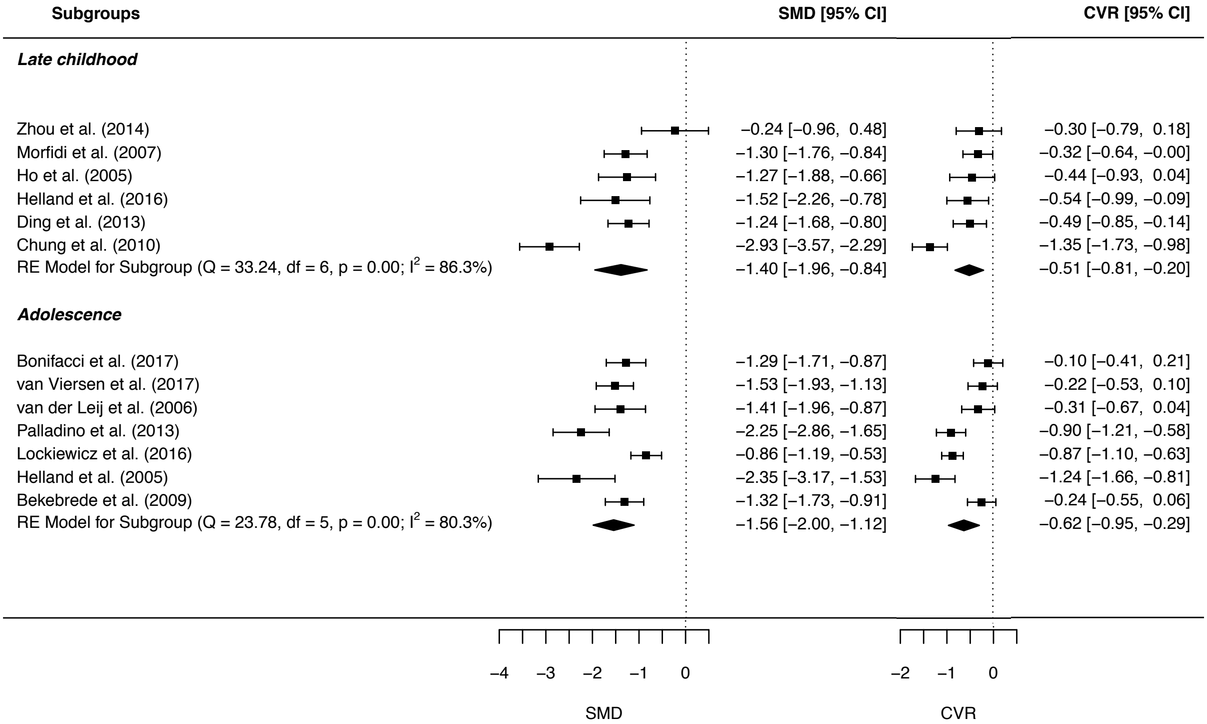
Figure 16. Moderator analysis on onset age of foreign language instruction for foreign language word reading



Note. SMD = Standardized Mean Difference*; CVR* = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

Age at foreign language assessment. Lastly, we compared the overall effects of a subgroup of studies that assessed participants in late childhood (age 6 to 11 - Bonifacci et al., 2017, Chung & Ho, 2010; Ding et al., 2013; Helland & Morken, 2016; Ho & Fong, 2005; Morfidi et al., 2007; Zhou et al., 2014) and adolescence (age 12 to 17 - Bekebrede et al., 2007; Helland & Kaasa, 2005; Lockiewicz & Jaskulskaa, 2016; Palladino et al., 2013; van der Leij & Morfidi, 2006; van Viersen et al., 2017). No significant differences between subgroups were found neither for the *SMD* or the *CVR* (*SMD*: $Z_{Diff^*} = 0.33$, $p = .74^{ns}$; *CVR*: $Z_{Diff^*} = 0.50$, $p = .61^{ns}$). Results are depicted in Figure 17.

Figure 17. Moderator analysis on age at foreign language assessment for foreign language word reading

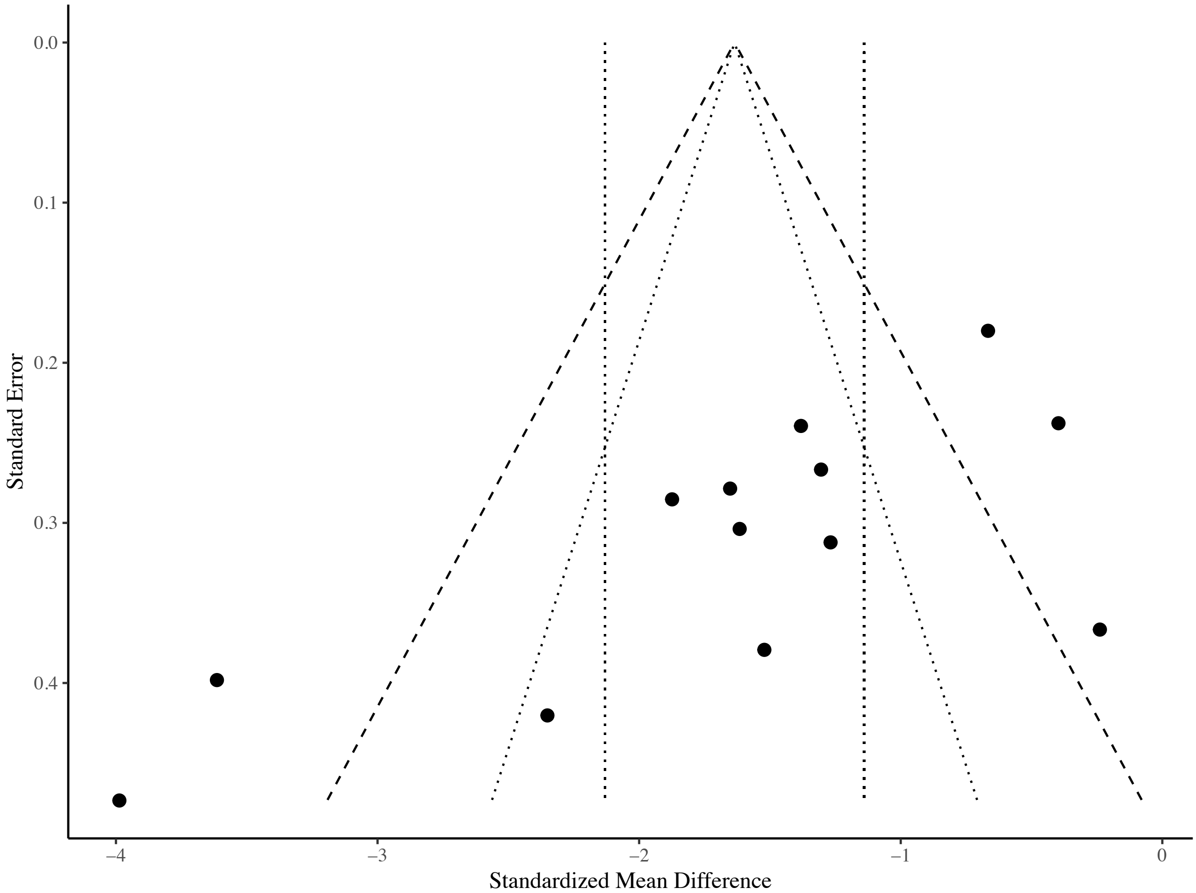


Note. SMD = Standardized Mean Difference*; CVR* = Natural logarithm of the ratio of coefficients of variation (Nakagawa et al., 2015).

Reporting bias

The influence of reporting bias could only be investigated with respect to the results of the meta-analysis on the *SMD*s for foreign language word reading due to limited data. Results of a funnel plot analysis with the trim and fill method by Duval and Tweedie (2000a, 2000b) and Duval (2005), as implemented in the *metafor* package in *R* (Viechtbauer, 2010), showed no evidence of reporting bias. The null hypothesis that the number of missing studies of the funnel plot was zero could not be rejected for the right (p = .25) or for the left side of the funnel plot (p = .50) (Viechtbauer, 2010). Figure 18 shows the corresponding funnel plot.

Figure 18. Funnel plot of studies measuring foreign language word reading (*n*=13)



Note. Each dot represents one study measuring foreign language word reading.

Sensitivity analysis

In order to assess if meta-analyses were influenced by our risk of bias assessment, we repeated the meta-analyses including the data reported by de Bree and Unsworth (2014), the only study that had been excluded due to a serious risk of bias in the selection of participants. This study contributed information to the outcome measures receptive vocabulary knowledge, word reading, nonword reading and orthographic knowledge. Results revealed a difference between the analysis with and without the data from de Bree and Unsworth (2014) for receptive vocabulary knowledge, but not for word reading, nonword reading and orthographic knowledge. For receptive vocabulary knowledge, the overall difference in average performance between groups was not significant without de Bree and Unsworth (2014), but reached significance when this study was included. The overall difference between performance variation between groups was not interpretable due to significant heterogeneity with and without de Bree and Unsworth (2014). Overall effects for word reading, nonword reading and orthographic knowledge could not be interpreted due to significant heterogeneity with and without de Bree and Unsworth (2014). Results are detailed in Appendix 4.

Discussion

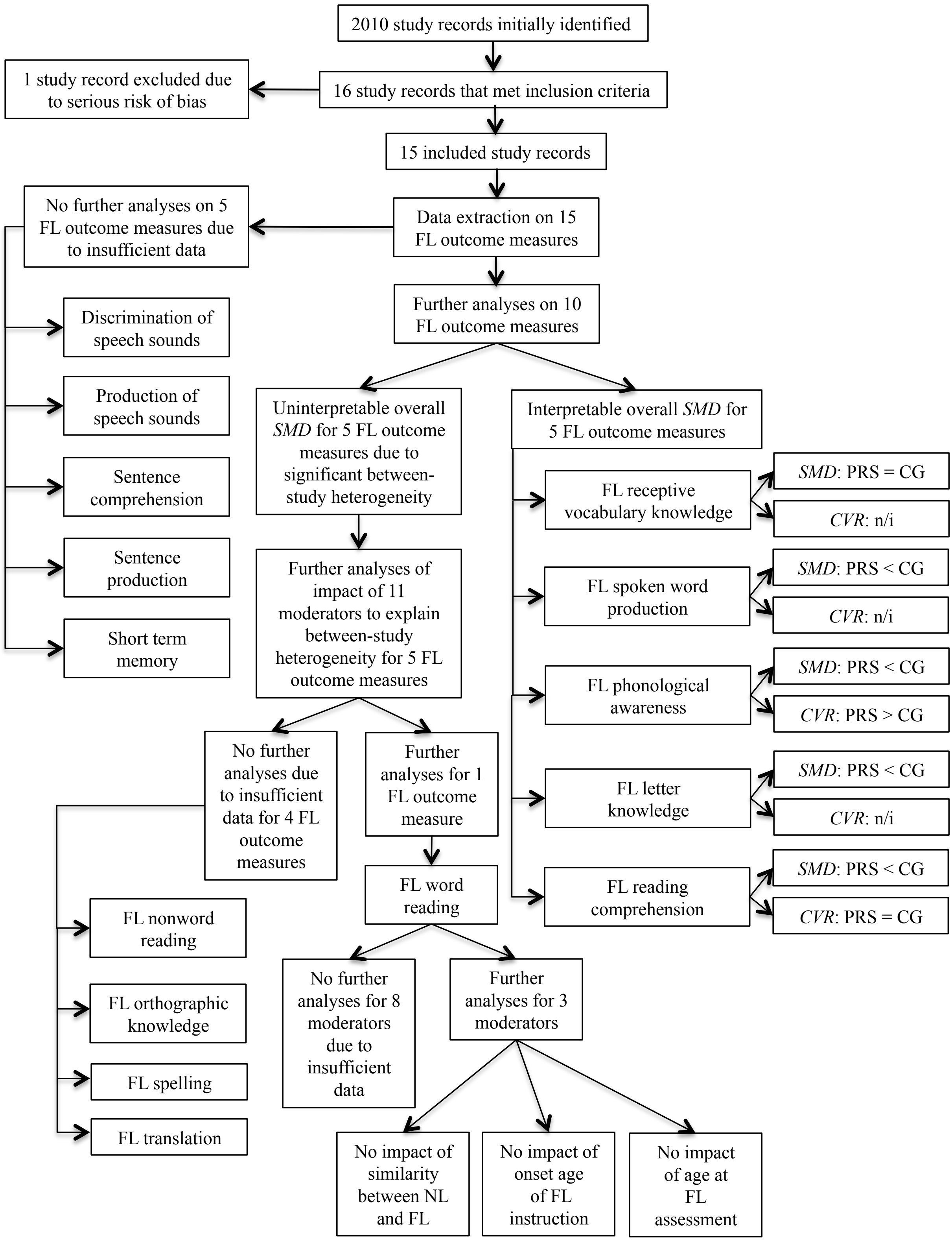
The structure of this section follows the guidelines provided by the Cochrane Collaboration (Higgins & Green, 2008). First, we present a summary of our main results, followed by an analysis of the overall completeness and applicability of the evidence that was summarized. Second, we assess the overall quality of the evidence for each foreign language outcome measure according to the Grades of Recommendation, Assessment, Development and Evaluation guidelines (GRADE, Schünemann et al., 2008). Lastly, we portray implications for practice and future research.

Summary of main results

Figure 19 summarises the main results of this review.

From a total of 2010 study records initially identified, only 16 (<1%) met inclusion criteria for the current review. Fifteen study reports displayed low to moderate risk of bias and were thus entered into the meta-analyses. Only one study was excluded due to serious risk of bias. For all the included studies, we extracted data on the 15 foreign language outcome measures considered in this review. Meta-analyses could not be conducted for the following five measures due to insufficient data: foreign language discrimination of speech sounds, production of speech sounds, sentence comprehension, sentence production and short term memory.

Figure 19. Summary of main results of this review



Note. FL = Foreign language. *SMD* = Overall standard men difference. *CVR* = Overall coefficient of variation. PRS = Poor readers/spellers. CG = Control group. NL = Native language. n/i = not interpretable due to significant between-study heterogeneity.

In contrast, we computed separate meta-analyses for the remaining ten foreign language outcome measures: receptive vocabulary knowledge, spoken word production, phonological awareness, letter knowledge, word reading, nonword reading, orthographic knowledge, reading comprehension, spelling and translation. Results on overall *SMDs* revealed significant between-study heterogeneity for word reading, nonword reading, orthographic knowledge, spelling and translation and so the interpretation of these values was not possible. To investigate the source of between-study heterogeneity, we planned to compute moderator analyses on 11 moderators that we assumed could have an impact on foreign language performance in children/adolescents with poor literacy skills. However, this was not possible for foreign language nonword reading, orthographic knowledge, spelling and translation due to insufficient data. We were able to conduct moderator analyses on foreign language word reading, although the information reported in the included studies was only enough to investigate the impact of three of the 11 moderators addressed in this review (i.e. similarity between native and foreign language, onset age of foreign language instruction and age at assessment). Information was insufficient for the remaining eight moderators. Results showed that neither the similarity between native and foreign language, onset age of foreign language instruction nor age at assessment could explain the between-study heterogeneity that we found for foreign language word reading.

For the five foreign language outcome measures for which meta-analytic results could be interpreted, overall *SMD*s showed that children/adolescents with poor literacy skills on average achieved a similar performance as the control group for foreign language receptive vocabulary knowledge. In contrast, their average performance was poorer on foreign language spoken word production, phonological awareness, letter knowledge and reading comprehension as compared to the control group. Complementary meta-analyses on the difference of performance variation (*CVR*) revealed that the performance of individual poor readers/spellers in phonological awareness varied significantly more than that of control participants. In contrast, performance in reading comprehension varied to a similar extent across both participant groups. Comparisons of overall performance variation between poor readers/spellers and control participants could not be determined for foreign language receptive vocabulary knowledge, spoken word production and letter knowledge, as *CVR* results reflected significant between-study heterogeneity.

Overall completeness and applicability of evidence

Overall completeness and applicability of the evidence summarized in this review is limited in three ways. First, the information summarized in this review only refers to some foreign language outcome measures (i.e., receptive vocabulary knowledge, word production, phonological awareness, letter knowledge, word reading, nonword reading, orthographic knowledge, reading comprehension, spelling and translation). In contrast, insufficient information was found for other foreign language outcome measures (i.e., foreign language discrimination and production of speech sounds, sentence comprehension and production skills). Hence, the findings are restricted to a small set of foreign language skills.

Second, foreign language attainment of children/adolescents with poor literacy skills was investigated with native speakers of a variety of Indo-European and non Indo-European native languages. However, information on some of the most widely spoken languages such as English and Spanish native speaking children/adolescents with poor literacy skills is currently missing. Furthermore, the foreign language investigated in all included studies was English. This limits the conclusions that can be drawn from this review with respect to other languages. Considering that the English orthography has been characterized as an 'outlier orthography' due to its irregular grapheme-phoneme mappings (Share, 2008) it is unclear how generalizable these results are to more 'regular' orthographies.

Lastly, and most importantly, significant heterogeneity between study effects seriously limited the interpretation of available evidence for several foreign language outcome measures (i.e. foreign language word reading, nonword reading, orthographic knowledge, spelling and translation). While most studies indicated a trend towards a lower foreign language attainment of children/adolescents with poor literacy skills as compared to the control group, the magnitude of this effect varied significantly from study to study. Where does this heterogeneity come from? In the registered protocol of this review, we anticipated 11 moderators that could represent potential sources of heterogeneity. Yet, due to limited data, we were only able to investigate the impact of three moderators (i.e. similarity between native and foreign language, onset age of foreign language instruction and age at assessment) on only one of 15 foreign language measures (i.e. foreign language word reading). While none of these three moderators contributed to explaining the observed between-study heterogeneity, it is possible that some of the other eight moderators, for which we did not have enough data, influenced results. For instance, the extent to which children/adolescents with poor and typical literacy skills differed in foreign language attainment might have varied if the study focused only on monolingual or bilingual participants (Nair et al., 2015; Tremblay & Sabourin, 2012). However, only five out of 16 studies provided information on this potential moderator. Similarly, different frequencies and durations of foreign language classes are likely to have influenced the results but could not be analysed due to insufficient data.

The lack of information on native language profiles of children/adolescents with poor literacy skills was especially alarming. Past research with children/adolescents with typical literacy skills has consistently reported associations between individual differences in native and foreign language attainment in children/adolescents with typical literacy skill (Cummins 1978, 1989, 2000; Dufva & Voeten, 1999; Meschyan & Hernández, 2002; Sparks & Ganschow, 1991; Sparks et al., 2006; Verhoeven, 1994). Although heterogeneous native language profiles of children/adolescents with poor literacy skills have been extensively described in the literature (e.g., Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013), only three of 16 studies included in this review did indeed distinguish between literacy profiles of poor readers/spellers in the selection of their participants. Therefore, studies included in this review most likely assessed children/adolescents with poor literacy skills with very variable native language profiles.

While some of the participants with poor literacy might have performed at a lower level than their peers with typical literacy skills on different foreign language outcome measures, others might have been just as successful as the control participants. The results on standardized mean differences *(SMDs*) between groups would mask this variability as they are based on group averages. In contrast, meta-analyses on differences of performance variation (*CVRs*) have recently been shown to be a useful tool to estimate the magnitude of within study variation (Nakagawa et al., 2015; Senior et al., 2016). We identified a clear pattern of higher performance variability in poor reader/speller participant groups than in control groups for the majority of foreign language outcome measures in most of the included studies in this review (see Figures 5-14). However, overall effects were predominantly not interpretable due to significant between-study heterogeneity. One potential source of between-study heterogeneity may indeed relate to the heterogeneous nature of poor reader/speller participants in the studies.

Quality of the evidence and potential biases in the review process

To assess the overall quality of the evidence for each of the foreign language outcome measures, we applied the Grades of Recommendation, Assessment, Development and Evaluation guidelines (GRADE, Schünemann et al., 2008). Following the guidelines, we began by judging all studies as "low quality of evidence", because they were all non-randomized trials. The evidence for each of the foreign language outcome measures could later be changed (i.e., upgraded or downgraded) following Schünemann et al.'s criteria (2008). We detail the specific reasons for each decision below.

For foreign language reading comprehension we upgrades the evidence to a level of moderate quality of evidence, due to the overall large *SMD* in the absence of significant heterogeneity between studies. Furthermore, children/adolescents with poor and typical literacy skills varied to a similar extent in their performance. In contrast, we maintained a judgement of low quality for the evidence on foreign language spoken word production, phonological awareness and letter knowledge. Despite obtaining large overall *SMDs* in the absence of significant between-study variance, it was unclear to what extent *SMDs* were representative of the performance of individual poor reader/speller participants. In the case of foreign language spoken word production and letter knowledge, the reasons for this were heterogeneous *CVR*s. In contrast, for foreign language phonological awareness, results showed a higher performance variation in the poor reader/speller participant group than in the control group.

Finally, we downgraded the quality of the evidence on foreign language receptive vocabulary knowledge, word reading, nonword reading, orthographic knowledge, spelling and translation to very low. For foreign language receptive vocabulary knowledge the results of a sensitivity analysis with and without the data of de Bree and Unsworth (2014), the only study that was excluded due to a serious risk of bias in the selection of participants, were inconsistent. Therefore, the evidence on foreign language receptive vocabulary knowledge in children/adolescents with poor literacy skills should be interpreted with caution. For foreign language word reading, nonword reading, orthographic knowledge, spelling and translation overall effects were not interpretable due to significant between-study variance. Moderator analyses for foreign language word reading did not contribute to explaining the observed variability and so results should also be interpreted with caution. No potential publication biases were identified.

Implications for practice

The results of this review provide evidence that children/adolescents with poor literacy skills on average show a lower attainment than their peers with typical literacy skills in foreign language spoken word production, phonological awareness, letter knowledge and reading comprehension. However, we also found evidence of higher performance variation in foreign language attainment of poor readers/spellers than of control participants. The sources of this variability have so far not been addressed by current research. Therefore, although children/adolescents with poor literacy skills seem to have a greater risk of experiencing foreign language learning difficulties, this might not be true for each individual child or adolescent with poor literacy skills. The common belief that children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills cannot be confirmed by the results of this review. Parents, teachers and clinicians should keep in mind that an individual student with poor literacy skills might be just as successful as other students with typical literacy skills. Instead of relying on the false common belief that all poor readers/spellers will struggle in learning a foreign language, foreign language attainment should be closely monitored and support and/or alternative teaching methods (e.g., less reliance on written work) put in place when necessary.

Implications for research

Available evidence on foreign language attainment in children/adolescents with poor literacy skills shows several limitations. Most importantly, future research should aim to better understand individual differences in foreign language attainment of children/adolescents with poor literacy skills.

First, an investigation of the impact of moderators related to participant characteristics might aid in better understanding the variability observed in foreign language attainment in children/adolescents with poor literacy skills. Past research has described heterogeneous native language profiles of children/adolescents with poor literacy skills that are likely to contribute to individual differences also in foreign language attainment (Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013; Moll & Landerl, 2009). Similarly, participants' linguistic background has been related to individual differences in foreign language attainment and is likely to be a source of the heterogeneity of foreign language attainment in poor readers/spellers reflected observed in the results of this review (Nair et al., 2015; Tremblay & Sabourin, 2012).

Second, information on the frequency, duration and onset age of foreign language instruction should be reported. This would make it possible to assess the impact of these variables in future meta-analyses. Furthermore, studies are needed to investigate foreign language attainment in children/adolescents with poor literacy skills learning a foreign language other than English. Although English is undoubtedly the most frequently instructed foreign language worldwide, we need to know if the difficulties observed in children/adolescents with poor literacy skills in learning English as a foreign language extend to other foreign languages. Related to this, future studies should also assess children/adolescents with poor and typical literacy skills that speak some of the most frequently spoken languages, such as English and Spanish, as native language. This would allow for a more representative picture of foreign language attainment in poor readers/spellers for a large amount of world´s population.

Third, while past research has focused on written foreign language attainment in children/adolescents with poor literacy skills, very little evidence is available on oral foreign language skills. This review aimed to collect information on eight foreign language measures. However, for five of these measures, insufficient information was available, as only one of the 16 studies included in this review reported data on these subskills (i.e. discrimination of speech sounds, production of speech sounds, sentence comprehension, sentence production and short term memory). The overall quality of the remaining three oral foreign language measures was judged as low and very low according to the GRADE guidelines (i.e. receptive vocabulary knowledge, spoken word production and phonological awareness). Future research should therefore measures oral foreign language attainment in children/adolescents with poor literacy skills. Especially for communicative purposes, understanding and being able to speak the foreign language seems just as, or even more important, than written foreign language performance.

Fourth, we can make several recommendations as to how future studies can decrease bias. In terms of selection of participants into the study, 15 out of 16 studies in this review either did not report how they selected participants into the study or explained that school staff (e.g. counsellors, teachers, etc.) had selected the participants. This is problematic because school staff could have selected participants based on their knowledge of the participants' foreign language performance. With respect to deviations from intended interventions, it is important to collect information on extra-curricular private foreign language tutoring, as Lockiewicz and Jaskulskaa (2016) did. It is possible that extra instruction contributes to explaining individual differences in foreign language attainment in children/adolescents with poor literacy skills. Lastly, future studies should report the presence or absence of missing data to make it possible for the reader to estimate the risk of bias. Only two out of 16 studies included in this review provided information in this respect.

Finally, future meta-analyses focusing on heterogeneous populations, such as children/adolescents with poor literacy skills, should include compute variation of outcome measures to capture individual differences. The common practice of only computing overall effect sizes based on central tendency measures such as the standardized mean difference (*SMD*) between groups can result in misleading answers to practically relevant research questions. For example, in the current review, only relying on overall *SMDs* would have led to the conclusion that children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills. However, the fact that foreign language performance varied more in children/adolescents with poor literacy skills than in control participants emphasizes that this conclusion might not apply to a significant proportion of poor reader/spellers.

References

*References marked with an asterisk indicate studies included in the systematic review.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language learning*, *59* (2), 249-306. doi: 10.1111/j.1467-9922.2009.00507.x

Barbiero, C., Lonciari, I., Montico, M., Monasta, L., Penge, R., Vio, C., Tressoldi, P.E., Ferluga, V., Bigoni, A., Tullio, A., Carrozzi, M. & Rofani, L. (2012). The submerged dyslexia iceberg: how many school children are not diagnosed? Results from an Italian Study. *PloS one*, *7*(10). doi: 10.1371/journal.pone.0048082

*Bekebrede, J., van der Leij, A., & Share, D. L. (2009). Dutch dyslexic adolescents: Phonological-core variable-orthographic differences. *Reading and Writing*, 22(2), 133-165. doi: 10.1007/s11145-007-9105-7

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second language research*, *13*(2), 116-137. doi:10.1191/026765897677670241

Bialystok, E. (2012). The impact of bilingualism on language and literacy development. In W.C. Ritchie, & T.K. Bhatia (Eds.) *The Handbook of Bilingualism and Multilingualism. Second Edition* (pp. 624-648). London: Blackwell Publishing.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*(01), 69-85. doi: 10.1017/S0142716400010584

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific studies of reading*, *9*(1), 43-61. doi: 10.1207/s1532799xssr0901_4

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

*Bonifacci, P., Canducci, E., Gravagna, G., & Palladino, P. (2017). English as a Foreign Language in Bilingual Language‐minority Children, Children with Dyslexia and Monolingual Typical Readers. *Dyslexia*, 23(2), 181-206. doi: 10.1002/dys.1553.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-analysis*. UK: John Wiley & Sons, Ltd.

British Dyslexia Association (2016). Modern Foreign Languages. Is it possible for someone who is dyslexic to learn a foreign language? Available at http://www.bdadyslexia.org.uk/educator/modern-foreign-languages [accessed 20/10/2016].

*Chung, K. K. H., & Ho, C. S. H. (2010). Second language learning difficulties in Chinese children with dyslexia: What are the reading-related cognitive skills that contribute to English and Chinese word reading? *Journal of Learning Disabilities*, 43(3), 195-211. doi: 10.1177/0022219409345018.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.) New York: Academic Press.

Coltheart, M. & Kohnen, S. (2012). Acquired and Developmental Disorders of Reading and Spelling. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp.892-920). New York, NY: Wiley. doi:10.1002/9781118432501.ch43

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press.

Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of cross-cultural psychology*, 9(2), 131-149.

Cummins, J. (1989). A theoretical framework for bilingual special education. *Exceptional children*, 56(2), 111-119.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. UK: Multilingual Matters Ltd.

*de Bree, E., & Unsworth, S. (2014). Dutch and English literacy and language outcomes of dyslexic students in regular and bilingual secondary education. Dutch Journal of *Applied Linguistics*, 3(1), 62-81. doi: 10.1075/dujal.3.1.04bre

De Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics*, 13, 1-24.

De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in second language acquisition*, 309-329.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(04), 499-533.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(03), 413-438. doi: 10.1017/S0142716410000056

*Ding, Y., Guo, J. P., Yang, L. Y., Zhang, D., Ning, H., & Richman, L. C. (2013). Rapid automatized naming and immediate memory functions in Chinese children who read English as a second language. *Journal of learning disabilities*, *46*(4), 347-362. doi: 10.1177/0022219411424209.

Dufva, M., & Voeten, M. J. (1999). Native language literacy and phonological memory as prerequisites for learning English as a foreign language. *Applied psycholinguistics*, 20(3), 329-348. doi: 10.1017/S014271649900301X.

Duval S.J., Tweedie, R.L. (2000a). A Nonparametric 'Trim and Fill' Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, 95(449), 89–98. doi: 10.1080/01621459.2000.10473905

Duval S.J., Tweedie, R.L. (2000b). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, 56(2), 455–463. doi: 10.1111/j.0006-341X.2000.00455.x

Duval, S. (2005). The trim and fill method. In H.R. Rothstein, A.J. Sutton, & M. Bornstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 128-144). West Sussex: John Wiley & Sons.

Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate*. New York: Cambridge University Press.

Francis, D., Caruana, N., Hudson, J. & McArthur, G. (2016). *The association between poor reading and anxiety and depression: a systematic review and meta-analysis protocol*. PROSPERO 2016:CRD42016049219. Available at http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016049219 [accessed 20/10/2016].

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, *99*(1), 529-534. doi: 10.1073/pnas.012611199

Friedmann, N., & Coltheart, M. (2016). Types of developmental dislexia. In A. Baron, & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives.* Berlin, Boston: De Gruyter Mouton.

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, *44*(9), 1146-1160. doi: 10.1016/j.cortex.2007.09.005

Friedmann, N., & Rahamim, E. (2007). Developmental letter position dyslexia. *Journal of Neuropsychology*, *1*(2), 201-236. doi: 10.1348/174866407X204227

Frost, R. (2012). A universal approach to modeling visual word recognition and reading: Not only possible, but also inevitable. *Behavioral and Brain Sciences, 35*(5), 310-329. doi:10.1017/S0140525X12000635

Goulandris, N. E. (2003). *Dyslexia in different languages: Cross-linguistic comparisons*. Whurr Publishers: London.

Haisma, J. (2009). Dyslexic Subtypes and Literacy Skills in L2 Opaque English: Met Name op het Gebied van het Leren van Verbaal Gedrag/taal. *Toegepaste Taalwetenschap in Artikelen*, *81*(1), 65-74. doi: 10.1075/ttwia.81.07hai

Hanley, J. R. (2017). Is There Just One Dyslexic Reader? Evidence for the Existence of Distinct Dyslexic Sub-Groups. *Current Developmental Disorders Reports*, *4*(4), 101-107. doi: 10.1007/s40474-017-0125-y

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

*Helland, T., & Morken, F. (2016). Neurocognitive development and predictors of L1 and L2 literacy skills in dyslexia: a longitudinal study of children 5–11 years old. *Dyslexia*, 22(1), 3-26. doi: 10.1002/dys.1515

Helland, T., Plante, E., & Hugdahl, K. (2011). Predicting dyslexia at age 11 from a risk index questionnaire at age 5. *Dyslexia*, 17(3), 207-226. doi: 10.1002/dys.432

Higgins, J.P., & Green, S. (2008). *Cochrane Handbook for systematic Reviews of Interventions: Cochrane Book Series*. Chichester, UK: John Wiley & Sons.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*, 327(7414), 557. doi: 10.1136/bmj.327.7414.557

*Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of psycholinguistic research*, 34(6), 603-618. doi: 10.1007/s10936-005-9166-1

International Dyslexia Association (2012). Definition of dyslexia. Available at https://dyslexiaida.org/definition-of-dyslexia/ [accessed 20/10/2016].

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99. doi: 10.1016/0010-0285(89)90003-0

Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in Grades 1–7. *Australian Journal of Learning Difficulties*, *20*(1), 15-38. doi: 10.1080/19404158.2015.1023209

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime'becomes 'smile': developmental letter position dyslexia in English. *Neuropsychologia*, *50*(14), 3681-3692. doi: 10.1016/j.neuropsychologia.2012.07.016

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex*. doi: 10.1016/j.cortex.2018.04.008

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false‐belief reasoning. *Developmental science*, *12*(1), 48-54. doi:10.1080/19404158.2015.1023209

Landerl, K., Wimmer, H., & Frith, U. (1996). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, 63(3), 315-334.

Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. USA: Oxford University Press.

*Łockiewicz, M., & Jaskulska, M. (2016). Difficulties of Polish students with dyslexia in reading and spelling in English as L2. Learning and Individual Differences, 51, 256-264. doi: 10.1016/j.lindif.2016.08.037

McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E., & Castles, A. (2013). Getting to grips with the heterogeneity of developmental dyslexia. *Cognitive neuropsychology*, *30*(1), 1-24. doi: 10.1080/02643294.2013.784192

McDougall, P., Borowsky, R., MacKinnon, G. E., & Hymel, S. (2005). Process dissociation of sight vocabulary and phonetic decoding in reading: A new perspective on surface and phonological dyslexias. *Brain and Language*, 92(2), 185-203. doi: 10.1016/j.bandl.2004.06.003

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. doi: 10.1037/a0033890

Meschyan, G., & Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology*, 94(1), 14-22. doi:10.1037/0022-0663.94.

Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*(5), 359-382. doi: 10.1080/10888430903162878

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., ... & Tóth, D. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65-77. doi: 10.1016/j.learninstruc.2013.09.003

*Morfidi, E., Van Der Leij, A., De Jong, P. F., Scheltinga, F., & Bekebrede, J. (2007). Reading in two orthographies: A cross-linguistic study of Dutch average and poor readers who learn English as a second language. *Reading and writing*, 20(8), 753-784. doi: 10.1007/s11145-006-9035-9

Müller, B., Ahnert, P., Burkhardt, J., Brauer, J., Czepezauer, I., Quente, E., Boltze, J., Wilcke, A. & Kirsten, H. (2014). Genetic risk variants for dyslexia on chromosome 18 in a German cohort. *Genes, Brain and Behavior*, *13*(3), 350-356. doi: 10.1111/gbb.12118.

Nair, V. K., Biedermann, B., & Nickels, L. (2016). Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism*, *20*(4), 473-487. doi: 10.1177/1367006914567005

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta‐analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143-152. doi: 10.1111/2041-210X.12309

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and cognitive processes*, *18*(4), 443-468. doi: 10.1080/01690960344000008

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Oxford, England: Cambridge University Press.

*Palladino, P., Bellagamba, I., Ferrari, M., & Cornoldi, C. (2013). Italian children with dyslexia are also poor in reading English words, but accurate in reading English pseudowords. *Dyslexia*, *19*(3), 165-177. doi: 10.1002/dys.1456

*Palladino, P., Cismondo, D., Ferrari, M., Ballagamba, I., & Cornoldi, C. (2016). L2 spelling errors in Italian children with dyslexia. *Dyslexia*, 22(2), 158-172. doi: 10.1002/dys.1522

Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, *37*(04), 813-840. doi: 10.1177/1362168816679030

Sampson, G. (2015). *Writing systems* (2nd ed.). Leeds: Equinox.

Schünemann, H.J. et al. (2008). Interpreting results and drawing conclusions. In J. PT Higgins & S. Green (Eds.) Cochrane Handbook for systematic Reviews of Interventions: Cochrane Book Series (pp.187-241). Chichester, UK: John Wiley & Sons.

Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, medicine, and public health*, *2016*(1), 244-255. doi: 10.1093/emph/eow020.

Siegel, L.S. (1988a). Definitional and theoretical issues and research on learning disabilities. *Journal of Learning Disabilities*, 21, 264-266.

Siegel, L.S. (2007). Perspectives on dyslexia. *Paediatrics & Child Health*, 11, 581-588.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of psychology*, *94*(2), 143-174. doi: 10.1348/000712603321661859

Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annu. Rev. Psychol.*, *59*, 451-475. doi: 10.1146/annurev.psych.59.103006.093633

Snowling, M. J., Duff, F., Petrou, A., Schiffeldrin, J., & Bailey, A. M. (2011). Identification of children at risk of dyslexia: the validity of teacher judgements using 'Phonic Phases'. *Journal of Research in Reading*, *34*(2), 157-170. doi: 10.1111/j.1467-9817.2011.01492.x.

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreign Language Annals*, 49(2), 252-270. doi: 10.1111/flan.12196.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal*, 75(1), 3-16. doi:10.2307/329830.

Sparks, R. L., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992a). Identifying native language deficits in high- and low-risk foreign language learners in high school. *Foreign Language Annals*, 25(5), 403.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia*, 56(1), 129-160. doi: 10.1007/s11881-006-0006-2.

Sparks, R., Ganschow, L., & Pohlman, J. (1989). Linguistic coding deficits in foreign language learners. *Annals of Dyslexia*, 39(1), 177-195. doi:10.1007/BF02656908.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992b). Test Comparisons among Students Identified as High-Risk, Low-Risk, and Learning Disabled in High School Foreign Language Courses. *The Modern Language Journal*, 76(2), 142-159. doi:10.2307/329768.

Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hróbjartsson A, Kirkham J, Jüni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schünemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF, Higgins JPT. *ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions*. Available at http://www.riskofbias.info [accessed 25/10/2016]

Sterne JAC, Higgins JPT, Elbers RG, Reeves BC and the development group for ROBINS-I. *Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance,* updated 12 October 2016. Available at http://www.riskofbias.info [accessed 25/10/2016].

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216. doi: 10.1016/j.cognition.2017.06.024

Tremblay, M. C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *The Journal of the Acoustical Society of America*, *132*(5), 3465-3474. doi: 10.1121/1.4756955

*van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of learning disabilities*, *39*(1), 74-90. doi: 10.1177/00222194060390010701

*van Viersen, S., de Bree, E. H., Verdam, M., Krikhaar, E., Maassen, B., van der Leij, A., & de Jong, P. F. (2017). Delayed early vocabulary development in children at family risk of dyslexia. *Journal of Speech, Language, and Hearing Research*, 60(4), 937-949. doi: 10.1044/2016_JSLHR-L-16-0031

Verhoeven, L. T. (1994). Transfer in bilingual development: The linguistic interdependence hypothesis revisited. *Language learning*, 44(3), 381-415. doi:10.1111/j.1467-1770.1994.tb01112.x.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. doi: 10.18637/jss.v036.i03

Wight, M.C.S. (2015). Students With Learning Disabilities in the Foreign Language Learning Environment and the Practice of Exemption. *Foreign Language Annals*, 48(1), 39-55. doi: 10.1111/flan.12122

Wright, W. E. (2013). Bilingual education. In W.C. Ritchie, & T.K. Bhatia (Eds.) *The Handbook of Bilingualism and Multilingualism. Second Edition* (pp. 624-648). London: Blackwell Publishing.

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. doi: 10.1177/0956797610363406

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of experimental child psychology*, *86*(3), 169-193. doi: 10.1016/S0022-0965(03)00139-5

*Zhou, Y., McBride-Chang, C., Law, A. B. Y., Li, T., Cheung, A. C. Y., Wong, A. M. Y., & Shu, H. (2014). Development of reading-related skills in Chinese and English among Hong Kong Chinese children with and without dyslexia. Journal of experimental child psychology, 122, 75-91. doi: 10.1016/j.jecp.2013.12.003

Acknowledgements

We would like to express our gratitude to Deanna Francis, Nicholas Badcock and John Elias for kindly sharing their experiences in the process of conducting a systematic review.

Appendices

Appendix 1. Signalling questions to guide the selection process

Signalling questions	Inclusion	Exclusion
1. Do the participants in the study attend primary schools from year 1 onwards or secondary schools?	Yes	No
2. Does the study include participants with poor literacy skills?	Yes	No
3. Do the students with poor literacy skills show a performance of at least 1 SD, 1 year or 1 grade below the expected level on any of the following measures: word/nonword reading accuracy, reading fluency, spelling?	Yes	No
4. Is the condition of average or below average literacy performance of all participants measured by a test as opposed to self- or teacher reports?	Yes	No
5. Are the poor readers/spellers compared to control participants on any foreign language measure?	Yes	No
6. Are the participants allocated to the control or experimental group based on their performance on foreign language tasks?	No	Yes
7. Do the participants have access to the target foreign language outside of the foreign language instruction context (except limited access to music, films or travel experience)?	No	Yes
8. Does the study measure the foreign language attainment of students with poor literacy skills (oral/ written language)?	Yes	No
9. Is the foreign language attainment being measured thought to be a result of a foreign language instruction received in a classroom context?	Yes	No

Note. One answer pointing to exclusion was enough to eliminate the study from the review process

Appendix 2. Information on moderators available in the studies included in this review.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
Participant characteristics																	
NL profile																	
Oral/written language deficits		?	?	?	?	?	?	?	?	?		?	?	?	?	?	14
Poor spoken NL and poor written NL																	
Average spoken NL and poor written NL	+										+						2
Reading/spelling deficits	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	16
Poor readers/good spellers																	
Good readers/poor spellers																	
Poor readers/poor spellers																	

Note. += allocated to this subgroup; ? = no information; NL = Native language

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
Participant characteristics																	
NL profile																	
Reading deficit subtype		?	?	?	?		?	?	?	?	?	?	?	?	?	?	14
Sublexical reading deficit						+											1
Lexical reading deficit	+					+											2
Mixed or other reading deficits[a]	+																1
Linguistic background	?			?		?	?	?		?		?	?	?	?	?	11
Monolinguals		+	+		+				+								4
Bilinguals																	0
Mono- & Bilinguals											+						1

Note. += allocated to this subgroup; ? = no information; NL = Native language; [a]recognizing letters, ordering letters and moving letters between words

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
FL instruction																	
Frequency of FL instruction		?	?	?	?	?	?	?	?	?		?	?		?	?	13
Less than 2 classes per week																	
Between 2-4 classes per week	+										+			+			3
More than 4 classes per week																	
Duration of FL classes		?	?	?	?	?	?	?	?	?		?	?	?	?	?	14
Less than 30 minutes																	
Between 30-60 minutes	+										+						2
More than 60 minutes																	

Note. += allocated to this subgroup; ? = no information; FL = Foreign language.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
FL instruction																	
Language pairing between NL and FL																	
Structural differences																	
NL Indo-European / FL Indo-European	+	+		+		+	+	+		+	+	+	+	+	+		12
NL Indo-European / FL non-Indo-European																	
NL non-Indo-European / FL Indo-European			+		+				+							+	4
NL non-Indo-European / FL non-Indo-European																	
Writing system differences																	
Alphabetic NL/ alphabetic FL	+	+		+		+	+	+		+	+	+	+	+	+		12
Alphabetic NL/ ideographic FL																	
Ideographic NL/ alphabetic FL			+		+				+							+	4

Note. + = allocated to this subgroup; ? = no information; NL = Native language; FL = Foreign language.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
FL instruction																	
Language pairing between NL and FL																	
Writing system differences																	
Ideographic NL/ ideographic FL																	
Orthographic regularity within alphabetic writing systems			n/a		n/a				n/a							n/a	4
NL regular / FL regular																	
NL irregular / FL regular																	
NL regular / FL irregular	+	+		+		+	+	+		+	+	+	+	+	+		12
NL irregular / FL irregular																	

Note. + = allocated to this subgroup; ? = no information; n/a = not applicable; [a]NL = Native language; FL = Foreign language.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
FL instruction																	
Onset age of FL instruction		?		?		?						?	?				
Early childhood: onset age before 6 years			+						+							+	3
Late childhood: onset age from 6-11 years					+		+	+		+	+						5
Adolescence: onset age from 12-17 years	+													+	+		3
Early adulthood: onset age from 18 years onwards																	0

Note. += allocated to this subgroup; ? = no information; FL = Foreign language.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree & Unsworth (2014)	Ding et al. (2013)	Haisma (2009)	Helland & Kaasa (2005)	Helland & Morken (2016)	Ho & Fong (2005)	Lockiewicz & Jaskulskaa (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij & Morfidi (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies with this information
FL assessment																	
Age at FL assessment																	
Early childhood: before 6 years of age																	0
Late childhood: from 6-11 years of age		+	+		+			+	+		+					+	7
Adolescence: from 12-17 years of age	+			+		+	+			+		+	+	+	+		9
Early adulthood: 18 years of age onwards																	0

Note. += allocated to this subgroup; FL = Foreign language.

Appendix 3. Foreign language outcome measures of included studies.

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree et al. (2014)	Ding et al. (2013)	Haisma (2009)	Helland et al. (2005)	Helland et al. (2016)	Ho et al. (2005)	Lockiewicz et al. (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij et al. (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies
Discrimination of speech sounds																	0
Production of speech sounds																	0
Receptive vocabulary knowledge	+			+					+		+			+		+	6
Spoken word production			+		+				+		+			+			5
Sentence comprehension							+										1
Sentence production							+										1
Short term memory									+								1
Phonological awareness			+						+		+			+			4
Letter knowledge			+								+						2
Word reading	+	+	+	+	+		+	+	+	+	+	+		+	+	+	14
Nonword reading	+	+		+						+	+	+		+			7

Note. += includes data on this foreign language outcome measure

	Bekebrede et al. (2009)	Bonifacci et al. (2017)	Chung et al. (2010)	de Bree et al. (2014)	Ding et al. (2013)	Haisma (2009)	Helland et al. (2005)	Helland et al. (2016)	Ho et al. (2005)	Lockiewicz et al. (2016)	Morfidi et al. (2007)	Palladino et al. (2013)	Palladino et al. (2016)	van der Leij et al. (2006)	van Viersen et al. (2017)	Zhou et al. (2014)	Total of studies
Orthographic knowledge	+		+	+		+					+			+	+		7
Reading comprehension		+			+						+			+			4
Spelling		+				+	+	+	+	+			+		+		8
Translation							+	+									2
Total of measures	2	3	3	2	3	2	□	□	6	2	5	1	1	5	2	2	

Note. += includes data on this foreign language outcome measure

Appendix 4. Results of sensitivity analysis.

	Mean difference								Variance difference							
	without de Bree & Unsworth (2014)				with de Bree & Unsworth (2014)				without de Bree & Unsworth (2014)				with de Bree & Unsworth (2014)			
	SMD	*Z*	*Q*	I^2	*SMD*	*Z*	*Q*	I^2	*CVR*	*Z*	*Q*	I^2	*CVR*	*Z*	*Q*	I^2
Receptive vocabulary knowledge	-0.47	-2.59^{ns}	7.38^{ns}	45.81%	-0.42	-2.84*	7.79^{ns}	35.88%	-0.28	-1.66^{ns}	15.01*	73.35%	-0.22	-1.54^{ns}	17.26*	71.03%
Word reading	-1.60	-6.47*	114.33*	89.50%	-1.60	-6.77*	117.40*	88.92%	-0.56	-5.08*	60.43*	80.14%	-0.58	-5.69*	63.26*	79.45%
Nonword reading	-0.98	-4.85*	18.46*	72.92%	-0.91	-5.10*	20.08*	70.12%	-0.22	-2.22^{ns}	11.20*	55.37%	-0.22	-2.58^{ns}	11.23^{ns}	46.57%
Orthographic knowledge	-1.40	-4.82*	30.18*	83.43%	-1.40	-5.56*	30.28*	80.18%	-0.69	-2.47^{ns}	77.96*	93.58%	-0.66	-2.74^{ns}	79.10*	92.41%

Note. SMD = Standard Mean Difference; *CVR* = natural logarithm of the ratio of coefficients of variance (Nakagawa et al., 2015). The sigificance level was $p < .05$ for the *Q* statistic and $p < .005$ for the *Z* statistic (Bonferroni correction for 10 comparisons). Cells marked in grey show overall effects that could not be interpreted due to the presence of significant between study heterogeneity.

Supplemental materials

Difference between registered protocol and final report

1. We revised the introduction section in order to improve overall clarity. Changes focused on wording and order of the presented information, as well as the inclusion of supporting references. No new information was added.

2. Foreign language outcome measures were considered primary outcomes of this review, instead of moderators, as presented in the protocol. The reason for this modification was that at the protocol stage we had planned to perform one overall meta-analysis including all foreign language outcome measures. During the review process we decided to complete separate meta-analyses for each foreign language outcome measure, to capture the available evidence in a more detailed way. Therefore, in the final report, the foreign language outcomes measures are presented as primary outcomes, instead of moderators, as detailed in the protocol.

3. We re-structured the classification of moderators presented in the protocol. The broad classification between moderators related to participant characteristics, foreign language instruction and foreign language assessment was maintained. However, onset age of foreign language instruction and language pairing between native and foreign language are now considered foreign language instruction moderators and not participant characteristics moderators. Furthermore, age at foreign language assessment is included as a foreign language assessment moderator in this final report. We decided to perform these changes to improve overall clarity of the structure of data extraction and analysis.

4. The foreign language outcome measure reading was broken down into the following four measures: (a) letter knowledge, (b) word reading, (c) nonword reading and (d) orthographic knowledge. With this change, we aimed to capture more detail with respect to the available evidence on foreign language reading subskills in children/adolescents with poor literacy skills.

5. We re-named the foreign language outcome measure "vocabulary knowledge" as "receptive vocabulary knowledge" to make the contrast with "spoken word production" clearer and avoid potential misunderstandings.

6. In addition to synthesizing available data through standardized mean differences (*SMD*), as described in the protocol, we decided to also complete meta-analyses on the natural logarithm of the ratio of coefficients of variation (*CVR*) between participant groups (Nakagawa et al., 2015). This allowed us to capture the average difference in the performance of both participant groups, but also to assess performance variation.

Search strategies

OVID databases, searched on 10th February 2017

1. literacy difficult$.ti. or literacy difficult$.ab.
2. decoding difficult$.ti. or decoding difficult$.ab.
3. reading difficult$.ti. or reading difficult$.ab.
4. spelling difficult$.ti. or spelling difficult$.ab.
5. literacy deficit.ti. or literacy deficit.ab.
6. decoding deficit.ti. or decoding deficit.ab.
7. reading deficit.ti. or reading deficit.ab.
8. spelling deficit.ti. or spelling deficit.ab.
9. literacy disorder.ti. or literacy disorder.ab.
10. decoding disorder.ti. or decoding disorder.ab.
11. spelling disorder.ti. or spelling disorder.ab.
12. literacy disabilit$.ti. or literacy disabilit$.ab.
13. decoding disabilit$.ti. or decoding disabilit$.ab.
14. reading disabilit$.ti. or reading disabilit$.ab.
15. spelling disabilit$.ti. or spelling disabilit$.ab.
16. literacy delay.ti. or literacy delay.ab.
17. decoding delay.ti. or decoding delay.ab.
18. reading delay.ti. or reading delay.ab.
19. spelling delay.ti. or spelling delay.ab.
20. literacy impairment.ti. or literacy impairment.ab.
21. decoding impairment.ti. or decoding impairment.ab.
22. reading impairment.ti. or reading impairment.ab.
23. spelling impairment.ti. or spelling impairment.ab.
24. literacy problem$.ti. or literacy problem$.ab.
25. decoding problem$.ti. or decoding problem$.ab.
26. reading problem$.ti. or reading problem$.ab.
27. spelling problem$.ti. or spelling problem$.ab.
28. literacy dysfunction.ti. or literacy dysfunction.ab.
29. decoding dysfunction.ti. or decoding dysfunction.ab.
30. reading dysfunction.ti. or reading dysfunction.ab.
31. spelling dysfunction.ti. or spelling dysfunction.ab.
32. dyslexi$.ti. or dyslexi$.ab.

33. dysgraphi$.ti. or dysgraphi$.ab.
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21
35. second language.ti. or second language.ab.
36. foreign language.ti. or foreign language.ab.
37. bilingua$.ti. or bilingua$.ab.
38. additional language.ti. or additional language.ab.
39. 35 or 36 or 37 or 38
40. 34 and 39

ProQuest databases, searched on 10th February 2017

(((("literacy difficult*" OR "decoding difficult*" OR "reading difficult*" OR "spelling difficult*") OR ("literacy deficit" OR "decoding deficit" OR "reading deficit" OR "spelling deficit") OR ("literacy disorder" OR "decoding disorder" OR "reading disorder" OR "spelling disorder") OR ("literacy disability" OR "decoding disability" OR "reading disability" OR "spelling disability") OR ("literacy delay" OR "decoding delay" OR "reading delay" OR "spelling delay") OR ("literacy impairment" OR "decoding impairment" OR "reading impairment" OR "spelling impairment") OR ("literacy problem" OR "decoding problem" OR "reading problem" OR "spelling problem") OR ("literacy dysfunction" OR "decoding dysfunction" OR "reading dysfunction" OR "spelling dysfunction")) OR (dyslexi* OR dysgraphi*)) AND ("second language" OR "foreign language" OR bilingua* OR "additional language"))

Wiley databases, searched on 26th February 2017

("literacy difficult*" OR "decoding difficult*" OR "reading difficult*" OR "spelling difficult*" OR "literacy deficit" OR "decoding deficit" OR "reading deficit" OR "spelling deficit" OR "literacy disorder" OR "decoding disorder" OR "reading disorder" OR "spelling disorder" OR "literacy disability" OR "decoding disability" OR "reading disability" OR "spelling disability" OR "literacy delay" OR "decoding delay" OR "reading delay" OR "spelling delay" OR "literacy impairment" OR "decoding impairment" OR "reading impairment" OR "spelling impairment" OR "literacy problem" OR "decoding problem" OR "reading problem" OR "spelling problem" OR "literacy dysfunction" OR "decoding dysfunction" OR "reading dysfunction" OR "spelling dysfunction" OR dyslexi* OR dysgraphi*) AND ("second language" OR "foreign language" OR bilingua* OR "additional language")

PubMED database, searched on 26th February 2017

((((second AND language) OR (foreign AND language)) OR (bilingua OR bilingual OR bilingual' OR bilingual's OR bilinguale OR bilingualism OR bilinguality OR bilinguality' OR bilingually OR bilinguals OR bilinguals')) OR (additional AND language)) AND ((((((((((((((((((((((((((((((((dysgraphia OR dysgraphia' OR dysgraphias OR dysgraphic OR dysgraphic' OR dysgraphics) OR (dyslexia OR dyslexia' OR dyslexia's OR dyslexiac OR dyslexiadysorthography OR dyslexial OR dyslexiaresponse OR dyslexias OR dyslexiax OR dyslexic OR dyslexic' OR dyslexic's OR dyslexicon OR dyslexics OR dyslexics' OR dyslexie OR dyslexies)) OR (spelling AND dysfunction)) OR (reading AND dysfunction)) OR (decoding AND dysfunction)) OR (literacy AND dysfunction)) OR (spelling AND problem)) OR (reading AND problem)) OR (decoding AND problem)) OR (literacy AND problem)) OR (spelling AND impairment)) OR (reading AND impairment)) OR (decoding AND impairment)) OR (literacy AND impairment)) OR (spelling AND delay)) OR (reading AND delay)) OR (decoding AND delay)) OR (literacy AND delay)) OR (spelling AND disability)) OR (reading AND disability)) OR (decoding AND disability)) OR (literacy AND disability)) OR (spelling AND disorder)) OR (reading AND disorder)) OR (decoding AND disorder)) OR (literacy AND disorder)) OR (spelling AND deficit)) OR (reading AND deficit)) OR (decoding AND deficit)) OR (literacy AND deficit)) OR (spelling difficulties OR spelling difficulty)) OR (reading difficulties OR reading difficulty)) OR (decoding difficulties OR decoding difficulty)) OR literacy difficulties)

Web of Science, searched on 26th February 2017

TOPIC: (("literacy difficult*" OR "decoding difficult*" OR "reading difficult*" OR "spelling difficult*" OR "literacy deficit" OR "decoding deficit" OR "reading deficit" OR "spelling deficit" OR "literacy disorder" OR "decoding disorder" OR "reading disorder" OR "spelling disorder" OR "literacy disability" OR "decoding disability" OR "reading disability" OR "spelling disability" OR "literacy delay" OR "decoding delay" OR "reading delay" OR "spelling delay" OR "literacy impairment" OR "decoding impairment" OR "reading impairment" OR "spelling impairment" OR "literacy problem" OR "decoding problem" OR "reading problem" OR "spelling problem" OR "literacy dysfunction" OR "decoding dysfunction" OR "reading dysfunction" OR "spelling dysfunction" OR dyslexi* OR dysgraphi*) AND ("second language" OR "foreign language" OR bilingua* OR "additional language"))

Timespan: All years. **Indexes:** SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.

Risk of bias assessment tool

ROBINS-I form
Adapted from Sterne et al. (2016)

Domain	Signalling question	Response option
Risk of bias due to confounding	1.1 Is there potential for confounding of the effect of intervention in this study?	Y/ PY/ PN/ N
	If N/PN to 1.1: the study can be considered to be at low risk of bias due to confounding and no further signalling questions need to be considered.	
	If Y/PY to 1.1: answer the following signalling questions	
	1.2 Did the authors use an appropriate analysis method that controlled for all the important confounding domains?	Y/ PY/ PN/ N/ NI
	1.3 Were confounding domains that were controlled for measured validly and reliably by the variables available in this study?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	no confounding expected	Low
	(i) confounding expected, all known important confounding domains appropriately measured and controlled for *and* (ii) reliability and validity of measurement of important domains were sufficient, such that we do not expect serious residual confounding.	Moderate
	(i) at least one known important domain was not appropriately measured, or not controlled for *or* (ii) reliability or validity of measurement of an important domain was low enough that we expect serious residual confounding.	Serious
	(i) confounding inherently not controllable or (ii) the use of negative controls strongly suggest unmeasured confounding.	Critical
	No information on whether confounding might be present.	No information

Note. Y= Yes; PY = Probably Yes; PN = Probably No; N = No; NI = No Information

Domain	Signalling question	Response option
Risk of bias in selection of participants into the study	2.1 Was selection of participants into the study (or into the analysis) based on participant characteristics observed after the start of intervention?	Y/ PY/ PN/ N/ NI
	2.2 Do start of follow-up and start of intervention coincide for most participants?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	(i) All participants who would have been eligible for the target trial were included in the study *and* (ii) for each participant, start of follow up and start of intervention coincided.	Low
	(i) Selection into the study may have been related to intervention and outcome *and* the authors used appropriate methods to adjust for the selection bias; *or* (ii) start of follow up and start of intervention do not coincide for all participants; *and* (a) the proportion of participants for which this was the case was too low to induce important bias; *or* (b) the authors used appropriate methods to adjust for the selection bias; *or* (c) the review authors are confident that the rate (hazard) ratio for the effect of intervention remains constant over time.	Moderate
	(i) Selection into the study was related (but not very strongly) to intervention and outcome; *and* this could not be adjusted for in analyses; *or* (ii) start of follow up and start of intervention do not coincide *and* a potentially important amount of follow-up time is missing from analyses; *and* the rate ratio is not constant over time.	Serious
	(i) Selection into the study was very strongly related to intervention and outcome; *and* this could not be adjusted for in analyses; *or* (ii) a substantial amount of follow-up time is likely to be missing from analyses; *and* the rate ratio is not constant over time.	Critical
	No information is reported about selection of participants into the study or whether start of follow up and start of intervention coincide.	No information

Domain	Signalling question	Response option
Risk of bias in classification of interventions	3.1 Were intervention groups clearly defined?	Y/ PY/ PN/ N/ NI
	3.2 Was the information used to define intervention groups recorded at the start of the intervention?	Y/ PY/ PN/ N/ NI
	3.3 Could classification of intervention status have been affected by knowledge of the outcome or risk of the outcome?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	(i) Intervention status is well defined; ***and*** (ii) intervention definition is based solely on information collected at the time of intervention.	Low
	(i) Intervention status is well defined; ***and*** (ii) some aspects of the assignments of interventions status were determined retrospectively.	Moderate
	(i) Intervention status is not well defined; ***or*** (ii) major aspects of the assignments of intervention status were determined in a way that could have been affected by knowledge of the outcome.	Serious
	(Unusual) An extremely high amount of misclassification of intervention status, e.g. because of unusually strong recall biases.	Critical
	No definition of intervention or no explanation of the source of information about intervention status is reported.	No information

Domain	Signalling question	Response option
Risk of bias due to deviations from intended interventions	4.1 Were there deviations from the intended intervention beyond what would be expected in usual practice?	Y/ PY/ PN/ N/ NI
	4.2 Were these deviations from intended intervention unbalanced between groups *and* likely to have affected the outcome?	Y/ PY/ PN/ N/ NI
	4.3 Were important co-interventions balanced across intervention groups?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	(i) Any deviations from intended intervention reflected usual practice; ***or*** (ii) any deviations from usual practice were unlikely to impact on the outcome. The important co-interventions were balanced across intervention groups, ***and*** there were no deviations from the intended interventions (in terms of implementation or adherence) that were likely to impact on the outcome.	Low
	(i) There were deviations from usual practice, but their impact on the outcome is expected to be slight; ***or*** (ii) the important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; ***and*** the analysis was appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-intervention) that were likely to impact the outcome.	Moderate
	(i) There were deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome. (ii) The important co-interventions were not balanced across intervention groups, or there were deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; ***and*** (iii) the analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-interventions) that were likely to impact on the outcome.	Serious

Domain	Signalling question	Response option
Risk of bias due to deviations from intended interventions	(i) There were substantial deviations from usual practice that were unbalanced between the intervention groups and likely to have affected the outcome. (ii) There were substantial imbalances in important co-interventions across intervention groups, or there were substantial deviations from the intended interventions (in terms of implementation and/or adherence) that were likely to impact on the outcome; ***and*** (iii) the analysis was not appropriate to estimate the effect of starting and adhering to intervention, allowing for deviations (in terms of implementation, adherence and co-interventions) that were likely to impact on the outcome.	Critical
	No information is reported on whether there is deviation from the intended intervention.	No information

Domain	**Signalling question**	**Response option**
Risk of bias due to missing data *Applicable to each outcome*	5.1 Were outcome data available for all, or nearly all participants (>90%)?	Y/ PY/ PN/ N/ NI
	5.2 Were participants excluded due to missing data on intervention status?	Y/ PY/ PN/ N/ NI
	5.3 Were participants excluded due to missing data on other variables needed for the analysis?	Y/ PY/ PN/ N/ NI
	5.4 Are the proportions of participants and reasons for missing data similar across interventions?	Y/ PY/ PN/ N/ NI
	5.5 Is there evidence that results were robust to the presence of missing data?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	(i) Data were reasonably complete; *or* (ii) proportions of and reasons for missing participants were similar across intervention groups; *or* (iii) the analysis addressed missing data and is likely to have removed any risk of bias.	Low
	(i) Proportions of and reasons for missing participants differ slightly across intervention groups; *and* (ii) the analysis is unlikely to have removed the risk of bias arising from the missing data.	Moderate
	(i) Proportions of missing participants differ substantially across interventions; or reasons for missingness differ substantially across interventions; *and* (ii) the analysis is unlikely to have removed the risk of bias arising from the missing data; *or* missing data were addressed inappropriately in the analysis; or the nature of the missing data means that the risk of bias cannot be removed through appropriate analysis.	Serious
	(i) (Unusual) There were critical differences between interventions in participants with missing data; *and* (ii) missing data were not, or could not, be addressed through appropriate analysis.	Critical
	No information is reported about missing data or the potential for data to be missing.	No information

Domain	**Signalling question**	**Response option**
Risk of bias in measurement of outcomes *Applicable to each outcome*	6.1 Could the outcome measure have been influenced by knowledge of the intervention received?	Y/ PY/ PN/ N/ NI
	6.2 Were outcome assessors aware of the intervention received by study participants?	Y/ PY/ PN/ N/ NI
	6.3 Were the methods of outcome assessment comparable across intervention groups?	Y/ PY/ PN/ N/ NI
	6.4 Were any systematic errors in measurement of the outcome related to intervention received?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	(i) The methods of outcome assessment were comparable across intervention groups; *and* (ii) the outcome measure was unlikely to be influenced by knowledge of the intervention received by study participants (i.e. is objective) or the outcome assessors were unaware of the intervention received by study participants; *and* (iii) any error in measuring the outcome is unrelated to intervention status.	Low
	(i) The methods of outcome assessment were comparable across intervention groups; *and* (ii) the outcome measure is only minimally influenced by knowledge of the intervention received by study participants; *and* (iii) any error in measuring the outcome is only minimally related to intervention status.	Moderate
	(i) The methods of outcome assessment were not comparable across intervention groups; *or* (ii) the outcome measure was subjective (i.e. vulnerable to influence by knowledge of the intervention received by study participants; *and* the outcome was assessed by assessors aware of the intervention received by study participants; *or* (iii) error in measuring the outcome was related to intervention status.	Serious
	The methods of outcome assessment were so different that they cannot reasonably be compared across intervention groups.	Critical
	No information is reported about the methods of outcome assessment.	No information

Domain	**Signalling question**	**Response option**
Risk of bias in selection of the reported result *Applicable to each outcome*	7.1 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple outcome *measurements* within the outcome domain?	Y/ PY/ PN/ N/ NI
	7.2 Is the reported effect estimate likely to be selected, on the basis of the results, from multiple *analyses* of the intervention-outcome relationship?	Y/ PY/ PN/ N/ NI
	7.3 Is the reported effect estimate likely to be selected, on the basis of the results, from different *subgroups*?	Y/ PY/ PN/ N/ NI
	Risk of bias judgement	
	There is clear evidence (usually through examination of a pre-registered protocol or statistical analysis plan) that all reported results correspond to all intended outcomes, analyses and sub cohorts.	Low
	(i) The outcome measurements and analyses are consistent with an *a priori* plan; or are clearly defined and both internally and externally consistent; *and* (ii) there is no indication of selection of the reported analysis among multiple analyses; *and* (iii) there is no indication of selection of the cohort or subgroups for analysis and reporting on the basis of the results.	Moderate
	(i) Outcomes are defined in different ways in the methods and results section, or in different publications of the study; *or* (ii) there is a high risk of selective reporting among multiple analyses; *or* (iii) the cohort or subgroup is selected from a larger study for analysis and appears to be reported on the basis of the results.	Serious
	(i) There is evidence or strong suspicion of selective reporting of results; *and* (ii) the unreported results are likely to be substantially different from the reported results.	Critical
	There is too little information to make a judgement (for example if only an abstract is available for the study).	No information

8

Data extraction form customized in Covidence

Foreign language attainment of children and adolescents with poor literacy skills: a systematic review protocol

Data extraction form in Covidence

Identification

Add Note

Study details

Sponsorship Source

Country

Setting

Poor readers/spellers recruited in schools or in clinical settings

Comments

Author's contact details

Author's name

Institution

Email

Address

Additional Identification data

Add another field

If we suspect that the study is a duplicate, we will add a corresponding note here.

1

Methods

Add Note

Design

Other | Group comparison

- Parallel group ← *These two default options will not be used.*
- Crossover

Additional Methods data

Add another field *If we are unsure, if the study design is a group comparison, we will detail this here.*

Population

Add Note

Inclusion criteria

Information on the children's school grade and school type will be detailed here.

Exclusion criteria

Group differences

The selection criteria for the experimental group will be detailed here. We will collect information on the literacy measure used to determine poor or average literacy skills, the instruments and cut-off points that were used. Furthermore, exclusion criteria such as sensory impairments or general developmental delays will be mentioned here.

Additional Population data

Add another field

Baseline Characteristics

Add Characteristic

2

Characteristic	Experimental group	Control group
age: M (SD)		
boys		
girls		
monolinguals		
bilinguals		
plurilinguals		
nonverbal reasoning		
L1 spoken language		
L1 word spelling		
L1 pseudoword spelling		
L1 word reading accuracy		
L1 nonword reading accuracy		
L1 word reading speed		
L1 pseudoword reading speed		
L1 reading comprehension		

Interventions *Information on the L1-L1 combination will be detailed as a note.* Add Note

Type your intervention name

Add Intervention

Experimental group

Control group

Description of Interventions

Add a descriptor

3

Characteristic	Experimental group	Control group
Age at FL onset		
Frequency of FL classes per week		
Duration of class in min		
Total amount of FL instruction		

Outcomes

Add Note

Add Outcome To add an outcome, you must have at least one intervention.

FL total words read

	Endpoint		
	mean	SD	N
EXPERIMENTAL GROUP			
CONTROL GROUP			

Add a timepoint

FL pseudowords read

	Endpoint		
	mean	SD	N
EXPERIMENTAL GROUP			
CONTROL GROUP			

Add a timepoint

As many outcome tables as necessary can be added here, according to the outcome measures presented in each study.

Reasons for excluding studies at full text screening phase

Authors	Exclusion reason
Borodkin et al. (2014)	Participants in post secondary education[a]
Letnes (2014)	Participants in post secondary education[a]
Ferrari et al. (2012)	Poor literacy skills not 1 SD, year or grade below expected level[b]
Jared et al. (2011)	Poor literacy skills not 1 SD, year or grade below expected level[b]
Kahn-Horwitz et al. (2006)	Poor literacy skills not 1 SD, year or grade below expected level[b]
Keung et al. (2009)	Poor literacy skills not 1 SD, year or grade below expected level[b]
Sparks et al. (1998)	Poor literacy skills not 1 SD, year or grade below expected level[b]
Farukh et al. (2016)	Literacy performance measure by self- or teacher reports[c]
Ghonsooly et al. (2010)	Literacy performance measure by self- or teacher reports[c]
Meng et al. (2016)	Literacy performance measure by self- or teacher reports[c]
Russak et al. (2015)	Literacy performance measure by self- or teacher reports[c]
vanDaal et al. (1999)	Literacy performance measure by self- or teacher reports[c]
Caglar-Ryeng et al. (2010)	No comparison with controls on foreign language measure[d]
Crombie et al. (1997)	No comparison with controls on foreign language measure[d]
Björn et al. (2013)	Allocation to participant group based on foreign language performance[e]
Kalindi et al. (2015)	Allocation to participant group based on foreign language performance[e]
McBride-Chang et al. (2012)	Allocation to participant group based on foreign language performance[e]
Pfenninger et al. (2015)	Allocation to participant group based on foreign language performance[e]
Alanis et al. (2005)	Access to foreign language outside of instruction context[f]
Geva et al. (1994)	Access to foreign language outside of instruction context[f]
Gottardo (2002)	Access to foreign language outside of instruction context[f]
Gottardo et al. (2008)	Access to foreign language outside of instruction context[f]
Gupta et al. (2007)	Access to foreign language outside of instruction context[f]
Guzman-Orth et al. (2013)	Access to foreign language outside of instruction context[f]
Haigh et al. (2011)	Access to foreign language outside of instruction context[f]
Hedman (2012)	Access to foreign language outside of instruction context[f]
Hutchinson et al. (2004)	Access to foreign language outside of instruction context[f]
Kieffer (2014)	Access to foreign language outside of instruction context[f]
Kline et al. (1972)	Access to foreign language outside of instruction context[f]
Lallier et al. (2014)	Access to foreign language outside of instruction context[f]
Lesaux et al. (2007)	Access to foreign language outside of instruction context[f]
Limbos (2006)	Access to foreign language outside of instruction context[f]
Lipka et al. (2007)	Access to foreign language outside of instruction context[f]
Willis (2002)	Access to foreign language outside of instruction context[f]
Kormos et al. (2010)	No assessment of oral or written foreign language performance[g]
Sauve (2009)	No assessment of oral or written foreign language performance[g]
Alanis et al. (2005)	Duplicated study report[h]
Bonifacci et al. (2017)	Duplicated study report[h]
Borodkin et al. (2014)	Duplicated study report[h]
Chung et al. (2010)	Duplicated study report[h]
Farukh et al. (2016)	Duplicated study report[h]
Farukh et al. (2016)	Duplicated study report[h]
Gupta et al. (2008)	Duplicated study report[h]

Authors	Exclusion reason
Helland (2008)	Duplicated study report[h]
Helland et al. (2016)	Duplicated study report[h]
Helland et al. (2016)	Duplicated study report[h]
Lipka et al. (2012)	Duplicated study report[h]
McBride-Chang et al. (2012)	Duplicated study report[h]
Pfenninger et al. (2016)	Duplicated study report[h]
Tong et al. (2017)	Duplicated study report[h]
van derLeij et al. (2006)	Duplicated study report[h]
Wiss (1993)	No acces to full text
Amner (1933)	No empirical study report
Dulude (2012)	No empirical study report
Finelli-Thomsen et al. (2012)	No empirical study report
Gonzales et al. (1981)	No empirical study report
Kovelman et al. (2016)	No empirical study report
Lodej (2016)	No empirical study report
Nijakowska (2009)	No empirical study report
Nijakowska (2010)	No empirical study report
Schneider (2009)	No empirical study report
AquinoAndersen et al. (2016)	Qualitative analysis
Ghazaleh et al. (2011)	Qualitative analysis
Lockhart-Pedersen (2013)	Qualitative analysis
Marogna (2013)	Qualitative analysis
Ni'mah (2016)	Qualitative analysis
Szaszkiewicz (2013)	Qualitative analysis
Gamper (2013)	Single case study
Valdois et al. (2014)	Single case study

Note. Studies are ordered according to selection signaling questions and other criteria. [a]Selection criteria signaling question 1 (see Appendix 1); [b-g]Selection criteria signaling questions 3-9 (see Appendix 1); [h]Exactly the same report or duplicated report from the same sample.

Chapter 2

Individual differences in foreign language attainment of children with poor literacy skills

This chapter has been prepared to be submitted as:

von Hagen, A., Kohnen, S. & Stadie, N. (to be submitted). Individual differences in foreign language attainment of children with poor literacy skills.

Abstract

Past research reports that children with poor literacy skills, as a group, show lower foreign language attainment than children with typical literacy skills. However, there is also considerable evidence that children with poor literacy are a heterogeneous group so that group averages could be masking individual differences. This study systematically examined to what extent group averages on the foreign language performance of children with poor literacy skills reflect individual performance profiles of poor readers/spellers as compared to children with typical literacy skills. Furthermore, we investigated if children with poor literacy skills show individual differences in the foreign language subskills, in which they struggle the most. We assessed German native speaking children with poor and typical literacy skills (n = 64) on broad range of carefully selected English foreign language tasks. In line with previous research we found that children with poor literacy skills, as a group, scored lower than controls on foreign language measures of spoken word comprehension, spoken word production, word and nonword reading and spelling. However, at the individual level, 18 of the 32 poor readers/spellers were just as successful as the control group on all measures. Moreover, the 14 poor readers/spellers that did show a lower foreign language attainment, differed with respect to the tasks that they struggled on.

Keywords: dyslexia; literacy skills; foreign language learning; bilingualism; second language learning

Introduction

While learning to read and write is a task that the majority of children master relatively quickly, about 3-10% of children worldwide lag behind their peers in acquiring literacy skills (Snowling, 2013). These difficulties cannot be explained by medical, emotional or neurological disorders, or insufficient literacy instruction (Elliott & Grigorenko, 2014; IDA, 2012) and have been referred to using different terms: developmental dyslexia and/or dysgraphia, specific reading and/or spelling difficulty, reading and/or spelling impairment, deficit or disability (Elliott & Grigorenko, 2014; Siegel, 1988a, 2007). In this study, we adopt the term 'children with poor literacy skills' or 'poor readers/spellers' to refer to children who show a below average performance (at least 1 SD, grade or year below the expected level) on at least one of the following literacy skills: inaccurate or slow word or nonword reading or spelling skills (Elliott & Grigorenko, 2014; IDA, 2012).

There is considerable evidence that children with poor literacy are a heterogeneous group. For instance, some poor readers/spellers only show difficulties in written language tasks, while others also struggle in oral language tasks (e.g. poor spoken vocabulary) (Bishop & Snowling, 2004; Catts, Adlof, Hogan & Weismer, 2005; McArthur & Castles, 2013; Wong, Kidd, Ho & Au, 2010). Furthermore, some children only fall behind their peers in reading, while others only struggle in spelling tasks and again others experience both types of difficulties (Moll & Landerl, 2009). The specific aspects of reading and spelling that are most difficult for poor readers/spellers (e.g. reading or spelling words or nonwords) have also been shown to vary from child to child (Coltheart & Kohnen, 2012; Friedmann & Coltheart, 2016; McArthur et al., 2013; Kohnen, Colenbrander, Krajenbrink & Nickels, 2015). There are also children with poor literacy skills that have additional learning difficulties, such as developmental dyscalculia or attentional deficit disorder (Germanò, Gagliano & Curatolo, 2010; Landerl & Moll, 2010; McArthur, Hogben, Edwards, Heath & Mengler, 2000; Moll, Göbel & Snowling, 2015). The term 'children with poor literacy skills' is therefore often said to be an 'umbrella term' to refer to children with very heterogeneous deficits (McArthur et al., 2013; Zoccolotti & Friedmann, 2010).

While past research has documented the above-mentioned heterogeneity in the cognitive profiles of children with poor literacy skills, in practice, these individual differences are often overlooked. First, children are not necessarily assessed on a comprehensive enough battery of tests to capture this heterogeneity. Test batteries used to test for reading and/or spelling difficulties are rarely specific enough to detect difficulties in the different subcomponents involved in reading and spelling. For example, in a recent study with Australian students in grades 1-7, it was found that 8.4% scored below the average range only on an irregular word spelling test, 7.8% scored below average only on a nonword spelling test and 5.7% scored below average on both tests (Kohnen, Colenbrander, Krajenbrinnk & Nickels, 2015). This indicates that if only word spelling had been assessed, a significant proportion of children would have appeared to have intact spelling skills, when in fact they struggled with translating sounds into spellings (as tested on the nonword spelling task).

Another way of how individual differences among poor readers/spellers are overlooked in practice, concerns the way teachers, clinicians and researchers decide if a child has poor or typical literacy skills. Reading is thought to be a skill that is distributed normally in the population (Shaywity, Escobar, Bennet, Fletcher & Makuch, 1992). However, in practice, whether or not someone receives a diagnosis or label of “poor reader” is determined on the basis of cut-off scores. Children who score below a certain cut-off on a reading or spelling test are identified as poor readers/spellers (and may thus receive the intervention/ special provisions/ etc.), while their classmates, even though they sometimes only differ by one score, are characterised as typical readers/spellers. While it is, of course, necessary to have criteria which determine who receives a diagnosis and related help, it is important to keep in mind that there are aspects of these criteria that are arbitrary. In addition to arbitrary cut-off scores, the tests used to assess for a difficulty do not always reflect the range of difficulties that have been reported in the literature.

Individual differences of children with poor literacy skills are also disregarded when it comes to school-based foreign language instruction. We refer to the specific case in which children receive explicit instruction in a foreign language other than their native language and only have access to this new language during foreign language classes (except through limited access to music, films computer games, etc. in the foreign language). Related situations, in which children speak their native language at home, but learn another language at school that is dominant in the country they live in, are not subject of this study (e.g. children of immigrant families or so-called 'English language learners' - August & Shanahan, 2017).

Parents, teachers and specialists working with poor readers/spellers often expect that the difficulties students have already experienced in their native language will transfer to the new language (Sparks, 2016). Thus, they are concerned that poor readers/spellers will be worse than their classmates with typical literacy skills in learning a foreign language (Baker, 1996; Genesee, 2015; Hussien, 2014; Palladino, Bellagamba, Ferrari & Cornoldi, 2013; Peer & Reid, 2014; Sparks, 2016; Wight, 2015). In some cases, students with poor literacy are even exempted from foreign language classes based on this common belief (Palladino et al., 2013; Wight, 2015). This has important consequences for children's future opportunities, as foreign language skills are crucial competencies which can be important, for example, when interviewing for a well-payed job. Furthermore, poor readers/spellers that are exempted from foreign language classes may have limited access to other cultures and may miss out on many of the cognitive advantages associated with foreign language learning (e.g. advantage on inhibitory control - Bialystok & Majumder, 1998; advantage in theory of mind development – Kovács, 2009; advantage on metalinguistic knowledge – Bialystok, 2012).

Surprisingly however, there is no strong evidence base to support the common belief that children with poor literacy skills show a lower attainment than their peers with typical literacy skills (Wight, 2015; Sparks, 2009, 2016; von Hagen, Kohnen & Stadie, submitted). While several studies have indeed reported lower foreign language attainment for children with poor literacy skills as compared to children with typical literacy skills, these studies have predominantly been based on group averages that disregard individual differences between poor readers/spellers (e.g. Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006). This is problematic, because there is ample evidence documenting that children with poor literacy skills show important individual differences in the success they achieve on different language tasks (Bishop & Snowling, 2004; Catts, Adlof, Hogan & Weismer, 2005; Coltheart & Kohnen, 2012; Friedmann & Coltheart, 2016; Kohnen, Colenbrander, Krajenbrink & Nickels, 2015; McArthur et al., 2013; Wong, Kidd, Ho & Au, 2010; Zoccolotti & Friedmann, 2010). It is therefore possible that not all children with poor literacy skills experience difficulties in learning a foreign language (Sparks, 2009).

For example, Ghonsooly and Javadian (2010) indicated a significantly lower performance of 13-year-old Farsi native speakers with poor literacy skills (n = 10), as compared to control participants (n = 10) on a series of tasks measuring English as a foreign language (i.e. phoneme deletion, rhyme detection, backward digit span, rapid picture naming, written word picture matching, word reading and spelling and reading comprehension). Similarly, Farukh and Vulchanova (2016) reported that 8-9 year old children with poor literacy skills in Urdu attending Urdu (n = 20) and English schools (n = 14) scored below matched control groups from the same schools (n = 18 and n = 14 respectively) on an English as a foreign language spoken sentence elicitation task.

Group averages are reported because they are thought to be a good characterisation of the performance of the majority of the children within the group. As such, the results above might lead to the conclusion that all, or at least the vast majority of children with poor literacy skills struggle in learning a foreign language. However, given that we know from previous research how much the performance of poor reader/spellers can vary across individual children, this variability could extend to their foreign language attainment. It is possible that foreign language difficulties occur only in some children with poor literacy skills, rather than affecting the majority of poor readers/spellers. In the extreme, these group averages may not even reflect the performance pattern of a single individual or in a less extreme scenario, fail to capture the performance of some of the individuals within the group (Kohnen & Nickels, 2015; Nugent, 2006; Smith & Little, 2018).

Indeed, Helland and Kaasa (2005) and van der Leij and Morfidi (2006) provide evidence that not all children with poor literacy skills struggle in learning a foreign language. In Helland and Kaasa's (2005) study, based on a sample of 12-year-old Norwegian native speakers, a subgroup of children with poor literacy skills and average foreign language listening comprehension skills (n = 10) achieved equivalent scores to the control group on foreign language tasks, such as foreign language daily conversation and story telling. Only children with poor literacy skills performing poorly in tasks focusing on foreign language listening comprehension (n = 10) scored significantly below the control group (n = 20) on all foreign language tasks. Similar results were reported by van der Leij and Morfidi (2006) with 14-year-old Dutch native speakers with poor literacy skills. A subgroup of children with poor literacy skills and good foreign language orthographic knowledge (n = 8) achieved equivalent scores to the control group (n = 25) on most foreign language tasks (i.e. foreign language spoken word and sentence-picture matching, semantic fluency, serial rapid naming, speeded word and nonword reading, text reading accuracy and comprehension tasks). Only the subgroup of children with poor literacy skills and poor foreign language orthographic knowledge (n = 8) showed a below-average performance on all foreign language measures.

Helland and Kaasa (2005) and van der Leij and Morfidi (2006) combined traditional group comparisons with subgroup analyses to capture individual differences in foreign language attainment of children with poor literacy skills. Their findings highlight that not all children with poor literacy skills show a lower foreign language attainment as compared to their peers with typical literacy skills. However, the extent to which individual children with poor literacy skills show foreign language difficulties still remains unknown.

One way to address this issue is to use methods that have been applied in single case studies and case series (Caramazza & Coltheart, 2006; Castles, Kohnen, Nickels & Brock, 2014; Crawford, Garthwaite & Porter, 2010; Crawford & Howell, 1998; Nugent, 2006). Statistics which allow careful comparisons at the level of the individual have been used successfully in cognitive neuropsychology to describe deficits in children with developmental cognitive disorders, including children with poor literacy skills (e.g., Barisic, Kohnen & Nickels, 2017; Friedmann & Rahamim, 2007; Friedmann & Lukov, 2008; Kohnen et al., 2012; Larsen, Kohnen, McArthur & Nickels, 2018). By focusing on individuals rather than groups of participants as the unit of empirical investigation, case studies are inherently in a better position than group studies to capture individual differences (Smith & Little, 2018). This is especially important in heterogeneous populations, such as in children with poor literacy skills (Bishop & Snowling, 2004; Friedmann & Coltheart, 2016; McArthur et al., 2013; Moll & Landerl, 2009; Ziegler et al., 2008).

Based on the heterogeneity described with respect to children with poor literacy skills, it is also likely that even when foreign language learning difficulties are present, the specific foreign language subskills that are affected differ from child to child. Similarly to achieving proficiency in a native language, mastering a foreign language involves learning different subskills, such as, for example, the ability to discriminate between different phonemes, to produce language-specific speech sounds, to comprehend spoken and written words and to write or say words. Past research on some of these measures is, however, limited. Therefore, we do not know if children with poor literacy skills show a lower performance on all foreign language subskills or if specific deficits can be identified. Furthermore, it remains unclear if individual poor readers/spellers present the same deficits or if the foreign language subskills that are affected vary between individual children.

In this study, we addressed the above-mentioned gaps by investigating the foreign language performance of children with poor literacy skills on the following eight measures: speech sound discrimination and production, spoken word comprehension and production, word reading and spelling and nonword reading and spelling. We selected these measures because they reflect different aspects of language processing that are required to complete different tasks, as specified in models of language processing (e.g. Ellis & Young, 1988).

With respect to speech sound discrimination and production, no previous study has investigated children. To the best of our knowledge, the only available evidence is a study with adult participants (Soroli et al., 2010). Soroli and colleagues (2010) reported that a group of adult French native speakers with poor literacy skills performed as well as adults with typical literacy skills (n = 15 in each group) when discriminating and repeating nonwords with Korean plosives. These results show that adults with poor literacy skills can be just as successful as their peers with typical literacy skills in discriminating and producing foreign language speech sounds. However, these findings may not apply to children, as previous studies have shown differences between child and adult foreign language learners (Abrahamson & Hyltenstam, 2009; Bialystok, 1997; De Keyser, 2000).

Regarding comprehension of spoken foreign language words in children with poor literacy skills, results have been inconsistent. For example, Ho and Fong (2005) reported significantly lower performance for a group of children with poor literacy skills as compared to the control group on a spoken word-picture matching task. In contrast, de Bree and Unsworth (2014) and Morfidi and colleagues (2007) reported a lack of group differences for children with poor versus typical literacy skills on foreign language spoken word comprehension. Individual differences between the participants with poor literacy skills involved in each study are likely to be responsible for some of these inconsistencies, as findings by van der Leij and Morfidi (2006) suggest (see above).

Concerning foreign language spoken word production, available evidence indicates that children with poor literacy skills tend to be worse than their peers with typical literacy skills in this foreign language subskill (De Bree & Unsworth, 2014; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006). Previous studies have used several tasks to tap into different aspects of the foreign language spoken word production process, such as for example serial rapid naming of numbers and objects and semantic fluency (Bekebrede et al., 2007; de Bree & Unsworth, 2014; Ghonsooly & Javadian, 2010; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006). In all of these tasks, children with poor literacy skills are reported to perform significantly poorer than their peers with typical literacy skills. Hence, foreign language spoken word production may often be difficult to acquire for children with poor literacy skills.

With respect to foreign language literacy skills, most studies report lower performance of children with poor literacy skills as compared to their peers with typical literacy skills (de Bree & Unsworth, 2014; Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Lockiewicz & Jaskulskaa, 2016). However, different results have also been reported, for example by Palladino et al. (2013). These authors found foreign language word reading, but not pseudoword reading deficits in a group of 12-year-old Italian children with poor literacy skills. They suggest that learning to read in a native language with a regular orthography such as Italian might strengthen children´s ability to convert graphemes and phonemes to the extent that even poor readers/spellers are proficient in this mechanism as compared to their peers with typical literacy skills. Nevertheless, it remains unclear to what extent this evidence reflects foreign language profiles of individual children with poor literacy skills.

The present study

The present study aimed to contribute towards a better understanding of foreign language attainment in children with poor literacy skills in two ways. First, we examined to what extent group averages reflect individual performance profiles of poor readers/spellers as compared to children with typical literacy skills. This is important, because past research is predominantly based on group results that average the performance of individual poor readers/spellers with very variable performances. Therefore, we do not know if this evidence is representative of the foreign language performance of individual students with poor literacy skills.

Second, we investigated if children with poor literacy skills show a lower performance than their peers with typical literacy skills on eight foreign language subskills (i.e. speech sound discrimination and production, spoken word comprehension and production, word reading and spelling and nonword reading and spelling). This is important, because detailed information on the difficulties that children with poor literacy skills might experience on specific foreign language subskills is limited. Furthermore, as mentioned in relation to the first aim of this study, available data is predominantly based on group averages that might not be representative for the performance of individual poor readers/spellers.

To address the two aims of this study, we measured eight English foreign language skills of 32 German native speaking children with poor literacy skills and 32 matched control participants with typical literacy skills. First, we conducted group analyses comparing the average performance of children with poor and typical literacy skills on eight foreign language measures. In this way we aimed to replicate previous studies and be able to compare our results to past research (de Bree & Unsworth, 2014; Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006).

As a second step, we completed single case statistical analyses comparing the performance of each individual poor reader/speller to the mean performance of the control group (Crawford et al., 2010; Crawford & Howell, 1998). This allowed us to we examine to what extent group level results reflect the individual foreign language performance profiles of poor readers/spellers as compared to children with typical literacy skills.

Method

Schools and Participants

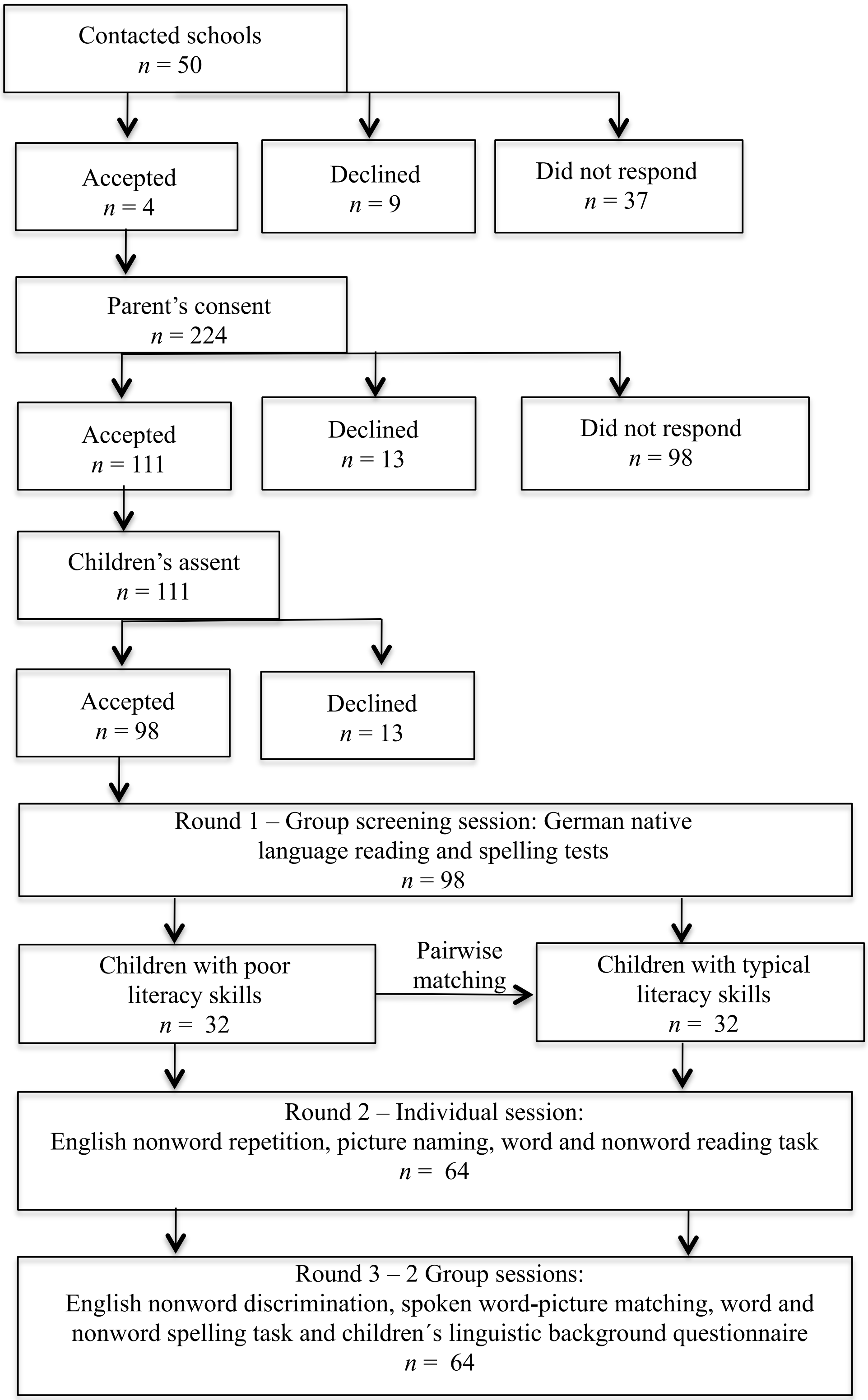
The complete data collection process is depicted in Figure 1.

Participating schools.

Fifty public primary schools in Berlin, Germany, were invited to participate in the project. Schools with specific teaching methodologies (e.g., Montessori: Montessori (2013), Waldorf: Steiner (1997), etc.) were excluded, as they were not representative of the typical education in German Primary schools. Four schools accepted the invitation.

All schools began to instruct English as a foreign language in grade 3, although some familiarized their students with English in an unsystematic playful way (e.g., by singing English Christmas songs) in grades 1 and 2. From grade 3 onwards, students across the four schools had the same amount of exposure to English, based on the regulations by the Primary school authority of Berlin. In grade 3, all children received two lessons of approximately 40-45 minutes per week. The amount of English instruction increased by one lesson a week resulting in five lessons a week in grade 6. None of the schools offered instruction in any other foreign language to the students who participated in this study. Although special accommodations (e.g. more time during examinations) for children with poor literacy skills can be solicited within the German school system, it is not possible to be exempted from foreign language classes during primary school.

Figure 1. Flowchart of the data collection process



The socio-economic status (SES) of the schools was assessed using the percentage of students exempted from paying an extra contribution for learning materials as an indicator (Müller, 2010). This measure served as an approximation of the school´s SES. Students are granted exemptions if their family receives financial support from the government due to a low family income. Half of the Primary schools of Berlin have an exemption rate of 38% or more of their students, a quarter of 53% or more and only 10 schools of 90% or more (Müller, 2010). In the current study, two of the participating schools were located in areas with a percentage lower than 40%, while the other two belonged to neighbourhoods in which up to 80% of the students are exempted from this fee.

Participating children.

After gaining ethics approval from the University of Potsdam and the education ministry of Berlin, we sent information letters, written consent forms and parent questionnaires to 224 parents of 11-12-year-old children in grade 6 of primary school students. One hundred and one parents agreed to their child's participation, but only 98 of these children gave their written assent and therefore, finally participated in the study (see Figure 1).

These students came from 10 different grade 6 classrooms and were taught by nine different English teachers. We selected this educational level, because students had been exposed to three complete years of English as a foreign language instruction. Therefore, we expected them to have gained sufficient foreign language knowledge to complete the eight foreign language tasks used in this study.

In order to assess which children met the inclusion and exclusion criteria for the present study, as a first step, parents completed a questionnaire on their children's developmental history. The exclusion criteria were uncorrected hearing or sight impairments and being bilingual with English as one of the native languages. None of the children who agreed to participate in the study showed any of these characteristics.

Next, the 98 children participated in a group screening session during which a German native language reading and spelling test were administered (see *Round 1* in Figure 1). Based on the results of this assessment, 32 children were assigned to the experimental group, which consisted of children with poor literacy skills. Children were allocated to this group if they scored in the lowest 20% on the standardised tests, either on the reading or the spelling screening test or on both. We chose this criterion based on the definition of poor readers specified in the manual of the ELFE, the standardized reading test that was used (Lenhard & Schneider, 2006). To form an equivalent control group, 32 further children who scored above the 20th percentile were matched pairwise to the experimental group on their learning environment (they were in the same class or instructed by the same teacher), their age, gender, and linguistic background (languages spoken at home).

The results from separate Fisher exact test analyses revealed no significant differences between the number of boys and girls in each group ($p > .05$) and the number of children that spoke only German at home and the children that spoke other languages at home ($p > .05$). The number of children attending schools in neighbourhoods in which less than 40% or up to 80% of parents are exempted from paying an extra contribution for learning materials ($p > .05$) was also equal across groups (Müller, 2010). Furthermore, we found no significant differences between the mean age of each group, as the results of a Mann Whitney *U* test showed, $U = 502.00$, $p = .89$, $r = .02$ (children with poor literacy skills: $M = 11.86$; $SD = 0.44$; $Mdn = 11.88$; children with typical literacy skills: $M = 11.82$; $SD = 0.32$; $Mdn = 11.75$). As expected due to the selection criteria, there was a significant between group difference with respect to the reading and spelling skills measured in the screening phase. Participants in the poor literacy group showed significantly lower scores as the control group on the reading screening test, $U = 126.00$, $p < .01$, $r = .65$ (children with poor literacy skills: $M = 29.50$; $SD = 5.00$; $Mdn = 30.50$; children with typical literacy skills: $M = 42.63$; $SD = 8.30$; $Mdn = 40.50$). The same pattern was observed with respect to the spelling screening test, $U = 159.00$,

$p < .01$, $r = .59$ (children with poor literacy skills: $M = 29.50$; $SD = 8.68$; $Mdn = 28.50$; children with typical literacy skills: $M = 40.03$; $SD = 5.00$; $Mdn = 39.50$).

Procedure

We completed three testing rounds (see Figure 1). Ninety-eight children participated in *Round 1*, which assessed native language reading and spelling skills to allocate children to the poor or typical literacy group. In the second and third testing round, we collected the foreign language experimental measures of 32 children with poor literacy skills and 32 pairwise matched children with typical literacy skills. The second testing round consisted of one individual session and the third testing round comprised two group sessions (max. of 12 children). All sessions had a duration of 45 minutes each. The instructions were given in the language the children were being tested in and the first author or a research assistant specifically trained for this project completed the assessment. Testing occurred in the second half of the school year, during regular school time, in a quiet room assigned by the school. Both testers were native German speakers with a proficient knowledge of English.

Materials

Questionnaires.

Parent questionnaire.

Parents were asked to complete a questionnaire with a multiple choice and brief answer format. This instrument provided information about the child's developmental history, their linguistic background and the presence of any learning difficulties.

Student questionnaire.

This instrument was added to the test battery after testing had started, because it became clear that some children had additional knowledge of languages which had not been mentioned by their parents. In a short questionnaire, we asked children about their linguistic background, their self-perception of their competencies in the languages they speak, and when they had started learning English. This information was used to complement the data provided by parents and to carefully match the experimental and control group.

Native language screening measures.

Reading test.

The *ELFE 1-6 Ein Lese-verständnistest für Erst- bis Sechstklässler* (Lenhard & Schneider, 2006) was used to assess German native language reading skills. This standardized group test is commonly used in Germany to measure reading skills at the word, sentence and text level. Although all three subtests were administered, only the results from the first subtest are relevant for the present study. In this subtest, participants were asked to choose the correct written target word out of four phonemically and graphemically similar distractor options that matched a picture. For example, they saw the picture of a window and had to mark *Fenster* (*window* in English) as the correct option. Distractor items were *Felsen* (*rock* in English), *Fehler* (*mistake* in English) and *Fremder* (*stranger* in English). Children were given two minutes to complete as many items as possible out of a total of 72. The internal consistency of the different subtests varies from $\alpha = .92$ to $\alpha = .97$ and the test-retest reliability measured after 14 days reveals a value of $rtt = .91$, with validity at $r = .71$ (Lenhard & Schneider, 2006). In this study, poor readers were defined as children who performed at or below the 20th percentile on the word level subtest.

Spelling test.

The *Hamburger Schreibprobe (HSP 5-10)* (May, 2012) was used to assess spelling ability. The test consists of a spelling to dictation task of 14 words, five sentences and a sentence cloze task. Internal consistency is reported between $r = .92$ and $r = .99$, and test-retest reliability is $rtt = .52$ to $rtt = .93$, depending on the retest timing; validity is reported as $r = .87$ (May, 2012). For the purpose of identifying poor spellers in this study, students who performed at or lower than the 20th percentile on the number of correctly written words score were selected.

Foreign language experimental measures.

Participants were assessed with bespoke tasks on the following foreign language measures: speech sound discrimination and production, receptive and expressive vocabulary, word and nonword reading, and word and nonword spelling. A description of all the experimental measures is provided in Table 1.

Table 1

Description of experimental measures

Foreign language measure	Task	N of items	Description
			All foreign language tasks were specifically designed considering the contrasts between German and English and taking into account the foreign language textbook vocabulary of the participants.
Speech sound discrimination	Nonword discrimination	26	Children listened to a pair of nonwords and were asked to mark on a response sheet if the pair was identical or not. The items contained only one English specific feature, which does not occur in German and filler phonemes, common to both languages. In addition to 20 pairs with English specific versus common phonological features, 6 more pairs were added. These items contrasted an English specific phoneme with a German specific phoneme. Half of the pairs were identical and half differed with respect to one phoneme. We selected the phoneme contrasts based on previous studies on FL sound discrimination (Bohn, Best, Avesani, & Vayra, 2011; Bohn & Flege, 1992; König & Gast; 2012; Trehub, 1976). See Appendix A for more details on the phonological contrasts that were used for all phonological measures and for the list of the items. The stimuli were recorded by a native speaker of German-English bilingual speaker with English as a dominant language. The internal consistency value corresponds to α = .29.
Speech sound production	Nonword repetition	24	Each child was asked to repeat a nonword that had been recorded by a native German-English bilingual speaker with English as a dominant language. Each nonword contained one English specific phoneme that does not exist in German (König & Gast; 2012). These target phonemes occurred once in a mono-, a bi- and a trisyllabic nonword respectively. The filler phonemes were common to both languages. The mean phoneme length of all items was 5.25 with a range of 3-8. See Appendix A for the list of items. Two native English speakers judged the native likeliness of the responses focusing on the target phoneme contrast. A double scoring of 22% of the recordings revealed an inter-rater reliability of κ=.77.
Spoken word comprehension	Spoken word-picture matching	48	48 items were selected from the FL textbooks used by the different schools from grade 3 to grade 6 (8 different books). We only used items present in all textbooks and marked as core vocabulary. The same items were used for the receptive and expressive tasks. We controlled for the following psycholinguistic variables, which have been shown to influence second language vocabulary learning: word class, time point of exposure during FL instruction, similarity to first language vocabulary (cognate status), presence of consonant clusters and syllable length (see Appendix B for details). For the receptive task children heard a word recorded by a German-English bilingual speaker and were asked to mark the correct option between four pictures on a response sheet. For the expressive task children saw two pictures and heard a sentence for the first picture. Their response to the second picture was elicited by an incomplete sentence (e.g. This is slow. This is……). The internal consistency value corresponds to α = .81 for the receptive task and to α = .84 for the expressive task.
Spoken word production	Picture naming	48	

Foreign language measure	Task	N of items	Description
Word reading	Word reading	24	We selected irregular words from the FL textbook vocabulary. Only words that were considered as irregular by the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) were included. Furthermore, the following psycholinguistic variables were controlled: time point of exposure during FL instruction, cognate status and presence of clusters (see Appendix C for list of items). Children were asked to read the words aloud. The internal consistency value corresponds to α = .84.
Nonword reading	Nonword reading	24	We designed a list of 24 monosyllabic nonwords including twelve grapheme-phoneme correspondences (GPC) that differ between German and English. Each contrast occurred in two nonwords respectively. Different types of GPCs contrasts between both languages were distinguished (see Appendix C). We asked children to read the items aloud, while their responses were recorded and transcribed after the testing session. The total score corresponds to the correctly identified target GPC. The internal consistency value corresponds to α = .75.
Word spelling	Word spelling	24	The same word and nonword lists as in the FL reading task were used. Children heard the items that were recorded by a native speaker through a loudspeaker and were asked to write them on a response sheet. To decide whether a GPC was scored as correct or incorrect, the following criteria, which have been previously used by Kohnen, Colenbrander, Krajenbrink & Nickels (2015), were implemented: only GPC with a type frequency of 20 per 7981 and a token frequency of at least 20000 out of all words in the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) were scored as correct. Once again, the total score corresponds to the correctly spelled target PGC. The internal consistency value corresponds to α = .84 for the word spelling task and to α = .56 for the nonword spelling task.
Nonword spelling	Nonword spelling	24	

Results

First, we present descriptive statistics and group level results comparing the foreign language performance of children with poor and typical literacy skills on eight foreign language measures. Second, we provide information on the individual performance pattern of each of the poor reader/speller participants as compared to the control group.

Group level comparisons

Assumptions for normality and equality of variance were not met and therefore, we performed non-parametric Mann-Whitney U tests. Furthermore, a Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) was applied to correct for multiple comparisons.

The results (detailed in Table 2) show that the group of children with poor literacy skills performed significantly below the control group on all foreign language measures, with the exception of the nonword discrimination and nonword repetition tasks. Although average performances on the nonword discrimination and repetition tasks were relatively high, we did not interpret this as ceiling performance, because skewness values were within an acceptable range (< -2.00, George & Mallery, 2010 - nonword discrimination: poor reader/speller group = -0.86; control group = -1.03; nonword repetition: poor reader/speller group = -0.25; control group = -0.76).

Table 2

Descriptive statistics and Mann-Whitney U results for eight foreign language tasks

Foreign language measure	Children with poor literacy skills (N = 32)				Children with typical literacy skills (N = 32)				Mann-Whitney U		
	M (SD)	*Mdn*	*Min*	*Max*	*M (SD)*	*Mdn*	*Min*	*Max*	*U*	*p*	*r*
Nonword discrimination (n = 26)	19.47 (1.79)	20.00	14.00	23.00	19.63 (1.71)	20.00	14.00	22.00	487	.73	.04
Nonword repetition (n = 24)	20.41 (1.73)	20.00	16.00	23.00	21.03 (2.32)	22.00	15.00	24.00	402	.13	.19
Spoken word-picture matching (n = 48)	42.19 (4.37)	44.00	28.00	48.00	44.44 (3.68)	45.50	32.00	48.00	329	<.01**	.31
Picture naming (n = 48)	34.03 (6.27)	34.50	10.00	42.00	38.94 (4.09)	39.50	28.00	46.00	253	<.001***	.44
Word reading (n = 24)	15.41 (4.83)	15.00	1.00	23.00	19.06 (2.98)	19.50	12.00	23.00	276	<.001***	.40
Nonword reading (n =24)	13.22 (3.68)	14.00	6.00	17.00	16.31 (3.41)	16.50	10.00	20.00	275	<.001***	.40
Word spelling (n =24)	12.13 (4.84)	12.00	1.00	22.00	17.50 (3.09)	18.00	11.00	24.00	178	<.001***	.56
Nonword spelling (n =24)	12.03 (2.92)	12.00	6.00	23.00	14.47 (2.70)	14.00	8.00	23.00	293	<.01**	.37

Note. The number of items for each measure is presented in parentheses. Results are expressed in raw scores.* p < .05, two tailed. ** p < .01, two tailed. *** p < .001, two tailed

Individual level comparisons

In order to investigate to what extent individual children with poor literacy skills revealed a lower foreign language attainment than their peers with typical literacy skills, we compared the raw scores of each poor reader/speller to the mean score of the control group by using a modified t-test procedure introduced by Crawford and colleagues (2010; Crawford & Howell,1998) and implemented in the program Singlims_ES.exe (see Table 3). The results show that none of the 32 children with poor literacy skills exhibited the same pattern of difficulties across the six foreign language measures, as reported for the group comparison. In fact, a total of 18 (56%) children with poor literacy skills did not differ from the control group at all (see poor literacy participant (PLP) 15-32 in Table 3). The remaining 14 (44%) children were significantly worse on a minimum of one, and up to a maximum of five foreign language measures (see PLP1-14 in Table 3). As shown in Table 3, there is a wide range of variability with respect to the strengths and weaknesses observed in the foreign language performance of the 14 children with deficits on at least one foreign language task.

Table 3

Modified t-test results of the foreign language performance of each child with poor literacy skills

Participant	Nonword discrimination	Nonword repetition	Spoken word - picture matching	Picture naming	Word reading	Nonword reading	Word spelling	Nonword spelling	Total deficits[c]
PLP1 [a]	-0.53 [b]	0.48	-2.39*	-6.76*	-5.59*	-2.36*	-5.08*	-0.68	5
PLP2	0.00	-1.92	-0.71	-2.52*	-3.37*	-2.36*	-2.91*	-2.71*	5
PLP3	1.74	-1.12	0.53	-3.02*	-3.05*	-2.93*	-1.98	-2.71*	4
PLP4	-1.14	-0.75	-0.97	-2.27*	-3.05*	-0.66	-3.84*	-2.31*	4
PLP5	0.00	0.85	-1.68	-2.77*	-3.37*	-2.07*	-4.15*	-0.68	4
PLP6	-3.41*	0.85	-3.89*	-2.08*	-0.19	-0.09	-1.67	0.94	3
PLP7	-0.53	0.05	-1.42	-1.59	0.13	-2.36*	-2.29*	-1.49	2
PLP8	0.00	-0.32	-2.20*	-1.34	-1.78	-1.51	-2.60*	-1.09	2
PLP9	1.14	-0.32	-1.23	-1.15	-2.10*	-0.94	-3.22*	-1.90	2
PLP10	-1.14	-0.75	-0.97	-0.65	-1.78	-0.66	-2.29*	-2.31*	2
PLP11	1.14	0.85	0.53	-3.21*	1.08	-2.64*	-1.67	-1.09	2
PLP12	0.00	0.05	-0.97	-1.59	-0.82	-0.66	-2.60*	-0.68	1
PLP13	0.61	-0.32	-0.45	-1.59	-1.78	-1.22	-3.22*	0.53	1
PLP14	-1.14	-0.32	-0.45	-0.4	-1.78	-0.37	-3.84*	-1.90	1
PLP15	0.00	0.05	0.27	-0.65	-1.14	-0.09	-0.44	-0.68	0
PLP16	0.00	-0.75	0.98	0.54	1.08	-0.09	-1.05	-1.49	0
PLP17	0.61	-1.12	0.72	-0.4	0.45	1.90	0.49	0.13	0
PLP18	1.14	-0.75	0.01	-1.15	-1.14	-1.51	-0.75	-1.09	0
PLP19	0.00	-1.12	0.72	0.72	0.45	0.76	0.80	0.13	0
PLP20	-0.53	-0.32	0.01	0.29	-0.82	-0.66	-1.36	-1.09	0
PLP21	1.14	0.48	0.01	0.04	-1.46	-0.66	-0.13	-0.28	0
PLP22	-0.53	-0.32	-1.42	0.04	1.08	-0.09	-1.05	-1.90	0
PLP23	0.61	0.85	0.53	-0.21	0.13	-0.37	-0.44	0.53	0
PLP24	0.00	-0.75	0.53	-0.4	0.45	-0.66	-0.13	-1.09	0
PLP25	0.00	0.05	0.27	0.72	0.13	-0.37	-1.98	0.94	0
PLP26	0.61	-0.75	0.27	-1.34	-1.14	0.76	-0.75	-0.68	0
PLP27	-1.67	0.05	0.01	-0.40	-1.14	-1.51	-1.67	-1.90	0
PLP28	0.61	0.85	0.53	-1.15	-1.14	-1.79	-1.05	-1.09	0
PLP29	0.61	-0.32	0.01	-0.9	0.76	-1.22	-0.75	-1.09	0
PLP30	-0.53	0.05	0.27	0.29	0.76	-0.09	1.73	2.15	0
PLP31	-1.14	0.48	-0.25	0.29	-1.78	-0.66	-0.44	-0.68	0
PLP32	-0.53	0.48	-0.71	-1.15	-0.51	-0.94	-1.98	0.94	0
Total[d]	1	0	3	7	6	6	11	4	

Note. [a]PLP = Poor literacy participant; [b]We report SinglimES t-values for each comparison; [c]Total foreign language deficits per participant; [d]Total of participants with deficits per foreign language measure *two tailed significant *t* values corresponding to $p < .05$ are shaded in grey.

Discussion

The aims of this study were (1) to investigate to what extent group averages on the foreign language performance of children with poor literacy skills reflect individual performance profiles of poor readers/spellers as compared to their peers with typical literacy skills, and (2) to examine which types of foreign language subskills (i.e. speech sound discrimination and production, spoken word comprehension and production, word reading and spelling and nonword reading and spelling) children with poor literacy skills have difficulties with. We will now discuss each of the research aims in turn.

First, we found more than half of the children with poor literacy skills to be just as successful as their peers with typical literacy skills on the eight foreign language measures assessed in this study. This demonstrates a complete mismatch between group and individual comparisons. While group comparisons indicated a lower foreign language attainment for children with poor literacy skills on six out of eight foreign language measures, this was not true for any of the 32 children with poor literacy. In fact, 18 out of 32 children with poor literacy skills were just as successful as the control group on all of the foreign language measures. Thus, only 14 children showed a below average performance on one or more of the foreign language tasks. This is the first report documenting that group averages can be completely misleading when investigating foreign language attainment in children with poor literacy. While previous studies (Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006) have highlighted the individual variability present in foreign language attainment of children with poor literacy skills, the present study is unique in emphasizing this fact through a combined analysis of group averages and individual comparisons. This approach seems to overcome the limitations of previous group comparison studies with respect to the applicability of the results to individual participants (Nugent, 2006; Smith & Little, 2018).

What may be the sources of the individual variability in foreign language attainment in children with poor literacy skills? Possible factors include the heterogeneous native language abilities of poor children with poor literacy skills (Bishop & Snowling, 2004; McArthur et al., 2013; Moll & Landerl, 2009; Ramus et al., 2013), the characteristics of the native language and foreign language orthography (Palladino et al., 2013), as well as foreign language motivation and teaching methods (Miller-Guron & Lundberg, 2000). However, this study does not address this question. While we ultimately need to be able to account for the reasons of the variability, this study simply aimed to provide a systematic documentation of the extent to which poor readers/spellers vary in their performance across different foreign language subskills. Carefully designed case series, selecting children based on their actual profiles in their foreign language skills, could be one way to investigate the sources of variability in future studies (Schwartz & Dell, 2010; Smith &Little, 2018).

The second aim of our study was to examine which different foreign language subskills (i.e. speech sound discrimination and production, spoken word comprehension and production, word reading and spelling and nonword reading and spelling) pose difficulties for children with poor literacy skills. At the group level, children with poor literacy skills performed below the control group on measures of foreign language spoken word comprehension, spoken word production, word and nonword reading and spelling. In contrast, they achieved similar scores to the control group in foreign language speech sound discrimination and production tasks. These results are partially consistent with previous findings.

Regarding foreign language speech sound discrimination and production, we found no evidence that children with poor literacy skills achieve a lower performance than their peers with typical literacy skills (see Soroli et al., 2010 for similar findings with adults). This is interesting, because children with poor literacy skills are often reported to struggle on phonological tasks in their native language (Ramus & Szenkovits, 2008; Wagner & Torgesen, 1987; but see Castles & Coltheart, 2004). However, previous studies have also shown that even when a phonological deficit is present, it does not necessarily apply to all phonological subskills (Ramus & Szenkovits, 2008). It may be then that the ability to discriminate and repeat foreign language speech sounds is not a difficulty generally experienced by children with poor literacy (Ramus & Szenkovits, 2008; Soroli et al., 2010). It is also possible that the poor readers/spellers in our sample simply did not have a phonological deficit, as not all children with poor literacy skills have difficulties in this area (see Castles & Friedmann, 2014).

With respect to foreign language spoken word comprehension our results are consistent with van der Leij and Morfidi (2006) in finding that some, but not all children with poor literacy skills show difficulties in this foreign language subskill. In our sample, only three children had poor foreign language spoken word comprehension skills. Van der Leij and Morfidi (2006) suggest foreign language orthographic knowledge as a source of these individual differences. Their results show that children with poor foreign language orthographic knowledge were more likely to be poor on spoken foreign language measures, including foreign language spoken word comprehension. When investigating if foreign language orthographic knowledge (indexed by foreign language word reading and spelling tasks) was associated with foreign language spoken word comprehension in our sample, we found that this pattern for one of the 32 poor readers/spellers in our sample (see Table 3). In contrast, ten poor readers/spellers showed average foreign language spoken word comprehension skills despite presenting with below-average foreign language orthographic knowledge, while one child showed the opposite pattern. Thus, in our sample, there is no reliable association between these skills for the majority of children, indicating that other factors need to be considered when trying to account for why foreign language spoken word comprehension performance is poor for some children with poor literacy skills.

Regarding foreign language spoken word production, our results show that, as a group, children with poor literacy skills show lower performance than children with typical literacy skills. This is consistent with previous research (de Bree & Unsworth, 2014; Ho & Fong, 2005; Morfidi et al., 2007; van der Leij & Morfidi, 2006). Spoken word production was also the spoken foreign language subskill in which most children with poor literacy skills struggled (seven out of 32 children). Possibly, foreign language spoken word production represents a specific weakness for many children with poor literacy skills. However, larger studies are required to understand prevalence and to investigate potential causes.

For foreign language word and nonword reading and spelling, children with poor literacy skills were also found to show a lower performance than their peers with typical literacy skills. Again, this is consistent with previous findings (de Bree & Unsworth, 2014; Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Helland & Kaasa, 2005; Lockiewicz & Jaskulskaa, 2016). At an individual level however, only 11 poor readers/spellers scored below the control group on these measures. More children struggled more with spelling foreign language words than reading words, reading nonwords and spelling nonwords. As previously suggested by Palladino et al. (2013), this could be indicating that poor readers/spellers who have learned to read in a native language with a regular orthography such as German are better in reading and spelling foreign language nonwords than words. This could be a consequence of ample experience with grapheme-phoneme and phoneme-grapheme conversion in their native language, which could then be transferred to the foreign language. However, it is also possible that the performance is specific to this sample. More research is needed to investigate this idea and examine cross-linguistic associations between native and foreign language literacy skills in poor readers/spellers.

Overall, the current study provides evidence that individual differences in the foreign language attainment of children with poor literacy skills need to be addressed more carefully. The documented divergence between the results obtained from group comparisons and individual poor reader/spellers exemplifies how misleading group averages can be. Tasking individual differences into account is especially important when assessing populations which are known to be heterogeneous. Thus, whenever conclusions are made at the group level, they should be verified at the individual level (Grice et al. 2017). Therefore, case series approaches that inherently take into account individual differences between participants may be more adequate to investigate foreign language attainment in children with poor literacy skills (Schwartz & Dell, 2010). Alternatively, group studies could be undertaken after carefully selecting participants for a specific type of deficit (e.g., poor native language word spelling).

The results of this study have important practical implications for the foreign language education of children with poor literacy skills. Our findings remind us to be cautious when drawing conclusions based on group averages when we are studying populations with heterogeneous profiles, such as children with poor literacy skills. It is important for parents and teachers to be conscious that the common belief that children with poor literacy skills show a lower foreign language attainment than their classmates with average literacy skills, does not hold for every child with poor literacy skills. In fact, in our sample it was not even true for the majority of the children with poor literacy skills. In addition, we found that even when children do struggle, the areas of difficulties are also quite diverse. In order to help parents and teachers better predict whether or not their student or child is likely to struggle (and thus put support in place), possible sources of individual differences need to be investigated.

References

Abrahamsson, N. & Hyltenstam, K. (2009). Age of Onset and Nativelikeness in a Second Language: Listener Perception Versus Linguistic Scrutiny. *Language Learning*, 59, 249–306. doi:10.1111/j.1467-9922.2009.00507.x

August, D., & Shanahan, T. (2017). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. London: Routledge.

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). *The Celex Lexical Database (CD-ROM).* Linguistic Data Consortium, University of Pennsylvania.

Baker, C. (1996). *Foundations of Bilingual Education and Bilingualism*. (2.ed.) UK: Multilingual Matters Ltd.

Barisic, K., Kohnen, S., & Nickels, L. (2017). Developmental graphemic buffer dysgraphia in English: A single case study. *Cognitive neuropsychology*, 34(3-4), 94-118. doi: 10.1080/02643294.2017.1359154

Bekebrede, J., van der Leij, A., & Share, D. L. (2007). Dutch dyslexic adolescents: phonological-core variable-orthographic differences. *Reading and Writing*, *22*(2), 133–165. doi: 10.1007/s11145-007-9105-7

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second language research*, *13*(2), 116-137. doi: 10.1191/026765897677670241

Bialystok, E. (2012). The impact of bilingualism on language and literacy development. In W.C. Ritchie, & T.K. Bhatia (Eds.) *The Handbook of Bilingualism and Multilingualism. Second Edition* (pp. 624-648). London: Blackwell Publishing.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*(01), 69-85. doi: 10.1017/S0142716400010584

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

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14* (2), 131-158.

Bohn, O. S., Best, C. T., Avesani, C., & Vayra, M. (2011). Perceiving through the lens of native phonetics: Italian and Danish listener's perception of English consonant contrasts. In W.-S. Lee & E. Zee (Eds). *Proceedings of the 17th International Congress of Phonetic Sciences* (pp.336–339). Hong Kong: Department of Chinese, Translational and Linguistics, City University of Hong Kong.

Caramazza, A., & Coltheart, M. (2006). Cognitive neuropsychology twenty years on. *Cognitive Neuropsychology*, *23*(1), 3-12. doi: 10.1080/02643290500443250

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77-111. doi: 10.1016/S0010-0277(03)00164-1

Castles, A., & Friedmann, N. (2014). Developmental dyslexia and the phonological deficit hypothesis. *Mind & Language*, *29*(3), 270-285. doi: 10.1111/mila.12050

Castles, A., Kohnen, S., Nickels, L., & Brock, J. (2014). Developmental disorders: What can be learned from cognitive neuropsychology? *Philosophical Transactions of the Royal Society - B*, 369(1634), 20130407. doi:10.1098/rstb.2013.0407

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378-1396. doi: 10.1044/1092-4388(2005/096)

Coltheart, M. & Kohnen, S. (2012). Acquired and Developmental Disorders of Reading and Spelling. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp.892-920). New York, NY: Wiley. doi:10.1002/9781118432501.ch43

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482-486. doi: 10.1076/clin.12.4.482.7241

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245-260. doi: 10.1080/02643294.2010.513967

de Bree, E., & Unsworth, S. (2014). Dutch and English literacy and language outcomes of dyslexic students in regular and bilingual secondary education. *Dutch Journal of Applied Linguistics*, *3*(1), 62-81. doi: doi:10.1075/dujal.3.1.04bre deBreeUnsworth_DUJAL_accepted.pdf

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(4), 499-533. doi: 10.1.1.198.8160&rep=rep1&type=pdf

Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate*. New York: Cambridge University Press.

Ellis, A.W. & Young, A.W. (1988) *Human cognitive neuropsychology*. Hove: Erlbaum.

Farukh, A., & Vulchanova, M. (2016). L1, quantity of exposure to L2 and reading disability as factors in L2 oral comprehension and production skills. *Learning and Individual Differences*, *50*, 221-233. doi: 10.1016/j.lindif.2016.07.013

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. In A. Bar-On, & D. Ravid (Eds.), *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistics Perspectives*. Berlin, Boston, MA: De Gruyter Mouton.

Friedmann, N., & Haddad-Hanna, M. (2012). Letter position dyslexia in Arabic: From form to position. *Behavioural Neurology*, *25*(3), 193-203. doi: 10.3233/BEN-2012-119004

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, 44(9), 1146-1160. doi: 10.1016/j.cortex.2007.09.005

Friedmann, N., & Rahamim, E. (2007). Developmental letter position dyslexia. *Journal of Neuropsychology*, 1(2), 201-236. doi: 10.1348/174866407X204227

Genesee, F. (2015). Myths About Early Childhood Bilingualism. *Canadian Psychology* 56(1), 6-15. doi: 10.1037/a0038599

George, D. & Mallery, P. (2010). *SPSS for windows step by step. A simple study guide and reference* (10. Baskı)

Germanò, E., Gagliano, A., & Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental neuropsychology*, *35*(5), 475-493. doi: 10.1080/87565641.2010.494748

Ghonsooly, B. & Javadian, M. (2010). An Examination of Developmental Dyslexia among Iranian EFL Second Graders. *International Journal of Language Studies*, 4(4*),* 333-356.

Grice, J., Barrett, P., Cota, L., Felix, C., Taylor, Z., Garner, S., ... & Vest, A. (2017). Four Bad Habits of Modern Psychologists. *Behavioral Sciences*, *7*(3), 53. doi: 10.3390/bs7030053.

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

Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of psycholinguistic research*, *34*(6), 603-618. doi: 10.1177/0022219409345018

Hussein, A. M. (2014). The effect of learning English (FL) on learning of Arabic literacy (NL) in the primary school. *International Education Studies*, *7*(3), 88-98.

International Dyslexia Association (2012). Definition of dyslexia. Available at https://dyslexiaida.org/definition-of-dyslexia/ [accessed 20/10/2016].

Kohnen, S., & Nickels, L. (2015). Methods of cognitive neuropsychology. In J.D. Wright (Ed.), *International Encyclopedia of Social and Behavioral Sciences* (2nd ed., pp. 90–94). Oxford, UK: Elsevier. doi: 10.1080/19404158.2015.1023209

Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in Grades 1–7. *Australian Journal of Learning Difficulties*, *20*(1), 15-38. doi: 10.1080/19404158.2015.1023209

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When ‘slime’becomes ‘smile’: developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681-3692. doi:10.1016/j.neuropsychologia.2012.07.016

Kohnert, K., Windsor, J., & Miller, R. (2004). Crossing borders: Recognition of Spanish words by English-speaking children with and without language impairment. *Applied Psycholinguistics*, *25*(4), 543-564. doi: 10.1017/S0142716404001262

König, E. & Gast, V. (2012). *Understanding English-German Contrasts*. Berlin: Erich Schmidt Verlag.

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental science*, *12*(1), 48-54. doi:10.1080/19404158.2015.1023209

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, *51*(3), 287-294. doi: 10.1111/j.1469-7610.2009.02164.x

Larsen, L., Kohnen, S., McArthur, G., & Nickels, L. (2018). An investigation of grapheme parsing and grapheme-phoneme knowledge in two children with dyslexia. *Reading and Writing*, 31(4), 991-1015. doi: 10.1007/s11145-018-9823-z

Lenhard, W. & Schneider, W. (2006). *ELFE 1-6: ein Leseverständnistest für Erst-bis Sechstklässler* [A reading comprehension test for 1st to 6th graders]. Göttingen: Hogrefe.

Łockiewicz, M., & Jaskulska, M. (2016). Difficulties of Polish students with dyslexia in reading and spelling in English as L2. *Learning and Individual Differences*, 51, 256-264. doi: 10.1016/j.lindif.2016.08.037

May, P. (2012). *Die Hamburger Schreib-Probe (HSP) [The Hamburger Writing Probe]*. Stuttgart: VPM.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(7), 869-874.

McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E. & Castles, A. (2013). Getting the grips with the heterogeneity of developmental dyslexia. *Cognitive Neuropsychology* 30(1), 1-24. doi: 10.1080/02643294.2013.784192

McArthur, G.M., & Castles, A. (2013). Phonological processing deficits in specific reading disability and specific language impairment: Same or different. *Journal of Research in Reading*, 36, 280-302. doi: 10.1111/j.1467-9817.2011.01503.x

Miller-Guron, L., & Lundberg, I. (2000). Dyslexia and second language reading : A second bite at the apple ? *Reading and Writing: An Interdisciplinary Journal 12*: 41–61. doi: 10.1023%2FA%3A1008009703641.pdf

Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, *13*(5), 359-382. doi: 10.1080/10888430903162878

Moll, K., Göbel, S. M., & Snowling, M. J. (2015). Basic number processing in children with specific learning disorders: Comorbidity of reading and mathematics disorders. *Child neuropsychology*, *21*(3), 399-417. doi: 10.1080/09297049.2014.899570

Montessori, M. (2013). *The Montessori Method*. New Brunswick: Transaction publishers.

Morfidi, E., Van Der Leij, A., De Jong, P. F., Scheltinga, F., & Bekebrede, J. (2007). Reading in two orthographies: A cross-linguistic study of Dutch average and poor readers who learn English as a second language. *Reading and writing*, *20*(8), 753-784. doi: 10.1007%2Fs11145-006-9035-9?LI=true

Müller, F. (2010). Verteilung von Armut im Primarbereich in Berlin. [Distribution of poverty in Primary schools in Berlin]. *Zeitschrift für Inklusion* [Journal for Inlcusion]. Retrieved from http://www.inklusion-online.net/index.php/inklusion-online/article/view/110/110

Nugent, W. (2006). Integrating Single-Case and Group-Comparison Designs for Evaluation Research. *Journal of Applied Behavioural Science*, 32 (2), 209-226. doi: 10.1177/0021886396322007

Palladino, P., Bellagamba, I., Ferrari, M. & Cornoldi, C. (2013). Italian Children with Dyslexia are also Poor in Reading English Words, but Accurate in Reading English Pseudowords. *Dyslexia* 19, 165-177. doi: 10.1002/dys.1456

Peer, L. & Reid, G. (Eds.) (2014). *Multilingualism, literacy & dyslexia*. UK: Taylor & Francis Group.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit?. *The Quarterly Journal of Experimental Psychology*, *61*(1), 129-141. doi: 10.1080/17470210701508822

Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model. *Brain*, *136*(2), 630-645. doi: 10.1093/brain/aws356

Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive neuropsychology*, *27*(6), 477-494. doi: 10.1080/02643294.2011.574111

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*(3), 145-150. doi: 10.1056/NEJM199201163260301

Siegel, L.S. (1988a). Definitional and theoretical issues and research on learning disabilities. *Journal of Learning Disabilities*, 21, 264-266.

Siegel, L.S. (2007). Perspectives on dyslexia. *Paediatrics & Child Health*, 11, 581-588.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic bulletin & review*, 1-19. doi: 10.3758/s13423-018-1451-8

Snowling, M. (2013). Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs,* 13(1), 7-14.

Soroli, E., Szenkovits, G., & Ramus, F. (2010). Exploring dyslexics' phonological deficit III: foreign speech perception and production. *Dyslexia*, *16*(4), 318-340. doi: 10.1002/dys.415

Sparks, R. L. (2009). If You Don't Know Where You're Going, You'll Wind up Somewhere Else: The Case of “Foreign Language Learning Disability”. *Foreign Language Annals*, *42*(1), 7-26. doi: 10.1111/j.1944-9720.2009.01005.x

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreign Language Annals*, 49(2), 252-270. doi: 10.1111/flan.12196

Steiner, R. (1997). *The essentials of education* (Vol. 18). Hudson: SteinerBooks.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development, 47*, 466-472.

van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of learning disabilities*, *39*(1), 74-90. doi: 10.1177/00222194060390010701

von Hagen, A. Kohnen, S. & Stadie, N. (submitted). Foreign language attainment of children/adolescents with poor literacy skills: a systematic review and meta-analysis.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, *101*(2), 192-212. doi: 10.1037/0033-2909.101.2.192

Wight, M.C.S. (2015). Students With Learning Disabilities in the Foreign Language Learning Environment and the Practice of Exemption. *Foreign language Annals*, 48(1), 39-55. doi: 10.1111/flan.12122

Wong, A. M. Y., Kidd, J. C., Ho, C. S. H., & Au, T. K. F. (2010). Characterizing the overlap between SLI and dyslexia in Chinese: The role of phonology and beyond. *Scientific Studies of Reading*, *14*(1), 30-57. doi: 10.1080/10888430903242043

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*(1), 151-178. doi: 10.1016/j.cognition.2007.09.004

Zoccolotti, P., & Friedmann, N. (2010). From dyslexia to dyslexias, from dysgraphia to dysgraphias, from a cause to causes: a look at current research on developmental dyslexia and dysgraphia. *Cortex*, *46*(10), 1211-1215.

Appendices

Appendix A

Nonword discrimination and repetition tasks.

Table A.1

Contrasts between phonological systems of German and English

	German	English
Language specific phoneme	ç , x	θ , ð
	ʁ	ɹ
	t͡s	w
	øː , œ	ɜː
	yː , Y	æ
Phonetic difference	final devoicing of voiced obstruents -d, -g, -b	velarized syllable-final -ɫ
	initial cluster ʃl- , ʃm- , ʃp-	initial cluster sl- , sm- , sp-
Phonotactic difference	initial cluster plosive + nasal kn-, gn-, pn-	Initial cluster consonant + glide mj-, lj-, fj-

Note. This table was adapted from König & Gast (2012)

Table A.2

Foreign language nonword discrimination task.

Item	Minimal pair	Item A	Item B	Foreign language specific target phoneme	Native and foreign language common contrast phoneme
1	yes	θiːp	tiːp	θ	t
2	no	θiːp	θiːp	θ	θ
3	yes	θɪm	fɪm	θ	f
4	no	θɪm	θɪm	θ	θ
5	yes	dʊθ	dʊs	θ	s
6	no	dʊθ	dʊθ	θ	θ
7	yes	ði:t	di:t	ð	d
8	no	ði:t	ði:t	ð	ð
9	yes	ðɪm	vɪm	ð	v
10	no	ðɪm	ðɪm	ð	ð
11	yes	ðʊk	zʊk	ð	z
12	no	ðʊk	ðʊk	ð	ð
13	yes	wɪf	vɪf	w	v
14	no	wɪf	wɪf	w	w
15	yes	wi:m	li:m	w	l
16	no	wi:m	wi:m	w	w
17	yes	wi:f	ji:f	w	j
18	no	wi:f	wi:f	w	w
19	yes	pæf	pɛf	æ	ɛ
20	no	pæf	pæf	æ	æ
				Foreign language specific target phoneme	Native language specific contrast phoneme
21		ɹʊp	ʁʊp	ɹ	ʁ
22		ɹʊp	ɹʊp	ɹ	ɹ
23		spɪf	ʃpɪf	sp-	ʃp-
24		spɪf	spɪf	sp-	sp-
25		nɜːk	nøːk	ɜː	øː
26		nɜːk	nɜːk	ɜː	ɜː

Note: The items were presented in a randomised order.

Table A.3

Foreign language nonword repetition task.

Item	Target phonological feature	Syllable length	IPA transcription
1		1	θi:p
2	θ and ð	2	bi:ðɪm
3		3	mɪbəθi:p
4		1	ɹɪʃ
5	ɹ	2	gi:ɹɪp
6		3	mɪbəɹɪk
7		1	wi:f
8	w	2	di:wif
9		3	bi:zəwɪm
10		1	nɜ:k
11	ɜ:	2	hi:lɜ:t
12		3	hi:fəgɜ:t
13		1	dæt
14	æ	2	hɪpæf
15		3	ti:dəfæp
16		1	pɪd
17	-d, -g, -b	2	mi:lɪg
18		3	pi:fədʊb
19		1	sli:ʃ
20	sp-, sl-, sm-	2	beespɪk
21		3	ʃi:dɪsmət
22		1	mjʊə
23	mj, lj, fj	2	hɪljʊə
24		3	ʃi:vəfjʊə

Note: The items were presented in a randomised order.

Appendix B

Foreign language spoken word-picture matching and picture naming tasks.

Table B.1

Foreign language vocabulary test items.

Item	Word class	Time of exposure	N of syllables	Presence of cluster	Cognate status	Item
1	Noun	early exposure (3. & 4. grade)	1 syllable	cluster	cognate	hand
2					noncognate	plate
3				no cluster	cognate	cow
4					noncognate	girl
5			2-3 syllables	cluster	cognate	apple
6					noncognate	flower
7				no cluster	cognate	banana
8					noncognate	window
9		late exposure (5. & 6. grade)	1 syllable	cluster	cognate	ghost
10					noncognate	spoon
11				no cluster	cognate	fire
12					noncognate	bag
13			2-3 syllables	cluster	cognate	elephant
14					noncognate	vegetables
15				no cluster	cognate	calendar
16					noncognate	cinema
17	Adjective	early exposure (3. & 4. grade)	1 syllable	cluster	cognate	blue
18					noncognate	small
19				no cluster	cognate	red
20					noncognate	sad
21			2-3 syllables	cluster	cognate	orange
22					noncognate	cloudy
23				no cluster	cognate	sunny
24					noncognate	yellow
25		late exposure (5. & 6. grade)	1 syllable	cluster	cognate	brown
26					noncognate	fast
27				no cluster	cognate	long
28					noncognate	wrong
29			2-3 syllables	cluster	cognate	hungry
30					noncognate	angry
31				no cluster	cognate	open
32					noncognate	empty

Item	Word class	Time of exposure	N of syllables	Presence of cluster	Cognate status	Item
33	Verb	early exposure (3. & 4. grade)	1-2 syllables	cluster	cognate	fly
34						help
35						sleep
36						swim
37					noncognate	draw
38						play
39						climb
40						travel
41		late exposure (5. & 6. grade)		no cluster	cognate	sing
42						learn
43						run
44						wash
45					noncognate	read
46						write
47						catch
48						choose

Table B.2

Coding scheme used to determine cognate status of foreign language vocabulary items following Kohnert, Windsor & Miller (2004, p.524).

Feature Overlap	Scoring	Example
Initial sound	3 = Same consonant	banana – Banane
(0-3 points)	2 = Same vowel	open - offen
	1= Similar sound (e.g., same sound class or one of consonant cluster)	swim - schwimmen
	0 = Complete mismatch between initial sounds	small – klein
Number of syllables	2 = Same number of syllables	cow - Kuh
(0-2 points)	1 = Different by only 1 syllable	fly - fliegen
	0 = Different by more than 1 syllable	choose – aussuchen
Consonants	3 = >70% consonant overlap	fire - Feuer
(0-3 points)	2 = 50-70% consonant overlap	red - rot
	1 = <50% consonant overlap	play - spielen
	0 = No consonant overlap	sad – traurig
Vowels	2 = ≥80% vowel overlap	elephant - Elefant
(0-2 points)	1 = 50-80% vowel overlap	blue - blau
	0 = no vowel overlap	cloudy - windig

Note. This coding scheme was developed following following Kohnert, Windsor & Miller (2004, p.524). Each English stimulus word was scored relative to its German translation equivalent on the four different features, for a maximum score of 10. The items that presented a score from 0-4 points were considered non-cognates, while 5-10 points qualified as a cognate status.

Appendix C

Foreign language reading and spelling tasks.

Table C.1

Foreign language irregular word reading and spelling

Item nr.	Time of exposure	Cognate status	Presence of cluster	Item
1	early exposure (3. & 4. grade)	cognate	cluster	friend
2				cold
3				month
4			no cluster	live
5				shoe
6				house
7		noncognate	cluster	climb
8				small
9				great
10			no cluster	put
11				head
12				eye
13	late exposure (5. & 6. grade)	cognate	cluster	break
14				group
15				find
16			no cluster	love
17				fall
18				come
19		noncognate	cluster	fast
20				ask
21				should
22			no cluster	key
23				choose
24				get

Table C.2

Foreign language nonword reading and spelling

		Reading				Spelling			
Item nr.	GPC[a]/PGC[b] contrast	Item	Expected pronunciation	English GPC	Conflicting German GPC	Item	Expected spelling[c]	English PGC	Conflicting German PGC
1	new FL phoneme – common NL/FL grapheme	lan	læn	<a> = /æ/	<a> = /a/	læn	lan	/æ/ =<a>	/æ/ = ∅
2		baf	bæf			bæf	baf		
3		mur	mɜ:r	<ur> = /ɜː/	<ur> = /uə/	mɜ:r	mur (mir, mer)	/ɜː/ = <ur>	/ɜː/ = ∅
4		hurm	hɜ:m			hɜ:m	hurm (hirm)		
5		thix	θ iks	<th> = /θ/	<th> = /t/	θ iks	thix (thicks, thiks)	/θ/ = <th>	/θ/ = ∅
6		theck	θ ek			θ ek	theck (thek)		
7		wip	wɪp	<w> = /w/	<w> = /v/	wɪp	wip (whip)	/w/ = <w>	/w/ = ∅
8		wiss	wɪs			wɪs	wiss (wis, whiss, whis)		
9	new FL grapheme – common NL/FL phoneme	shim	ʃɪm	<sh> = /ʃ/	<sh> = ∅	ʃɪm	shim	/ʃ/ = <sh>	/ʃ/ = <sch>
10		shep	ʃep			ʃep	shep		
11	common NL/FL phoneme and grapheme, but new correspondence	noom	nu:m	<oo> = /uː/	<oo> = /oː/	nu:m	noom	/uː/ = <oo>	/uː/ = <u>
12		doot	du:t			du:t	doot		
13		knep	nep	<n> = /kn/	<kn> = /kn/	nep	knep (nep)	/n/ = <n>	/n/ = <n>
14		knoff	nɔf			nɔf	knoff (nof, knof, noff)		
15		bleen	bli:n	<ee> = /iː/	<ee> = /eː/	bli:n	bleen (blean, blene)	/iː/ = <ee>	/iː/ = <ie>
16		deeks	di:ks			di:ks	deeks (decks, deaks, deacks, deex, deax)		
17		jeck	ʤek	<j> = /ʤ/	<j> = /j/	ʤek	jeck (jek, geck, gek)	/ʤ/ = <j>	/ʤ/ = <dsch>
18		jit	ʤɪt			ʤɪt	jit (git)		
19		zill	zɪl	<z> = /z/	<z> = /ʦ/	zɪl	zill	/z/ = <z>	/z/ = <s>
20		zem	zem			zem	zem		
21		chim	ʧɪm	<ch> = /ʧ/	<ch> = /ç/	ʧɪm	chim	/ʧ/ = <ch>	/ʧ/ = <tsch>
22		chet	ʧet			ʧet	chet		
23		yoll	jɔl	<y> = /j/	<y> = /i/	jɔl	yoll (yol)	/j/ = <y>	/j/ = <j>
24		yem	jem			jem	yem		

Note. [a]GPC= Grapheme-phoneme correspondence; [b]Phoneme-grapheme correspondence; [c]We listed spellings other than the target expected spelling in brackets. These alternatives were accepted as correct based on the following criteria previously used by Kohnen, Colenbrander, Krajenbrink & Nickels (2015): only PGC with a type frequency of 20 per 7981 and a token frequency of at least 20000 out of all words in the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) were scored as correct

Chapter 3

Sources of individual differences in foreign language attainment of children with poor literacy skills

This chapter is currently under review as:

von Hagen, A., Stadie, N., Robidoux, S. & Kohnen, S. (submitted). Sources of individual differences in foreign language attainment of children with poor literacy skills

Abstract

Purpose: The purpose of this study was to investigate native language skills as a source of individual differences in foreign language attainment of children with poor literacy skills. We also investigated the influence of broader cognitive abilities (i.e. linguistic background, intellectual ability, short term and working memory and foreign language learning motivation) on foreign language attainment in poor readers/spellers.

Method: A sample of 32 German native speaking children with poor literacy skills (11-12 years), who were learning English as a foreign language, completed eight native and equivalent foreign language tasks (i.e. nonword discrimination and repetition, spoken word-picture matching, picture naming, nonword reading and spelling, word reading and spelling). In addition, we collected measures on children's linguistic background, intellectual ability, short term and working memory and foreign language learning motivation.

Results: Native language speech sound perception, spoken word production, nonword reading, word reading and spelling skills significantly contributed to explaining individual differences in equivalent foreign language measures. Furthermore, only foreign language learning motivation, but none of the other broader cognitive measures played a significant role in accounting for the observed variance in poor reader/spellers' foreign language performance.

Conclusion: Our findings underline the need to take into account native language skills to reach a better understanding of individual differences in foreign language attainment of children with poor literacy skills.

Keywords: dyslexia, literacy skills, foreign language learning, bilingualism, second language learning

Introduction

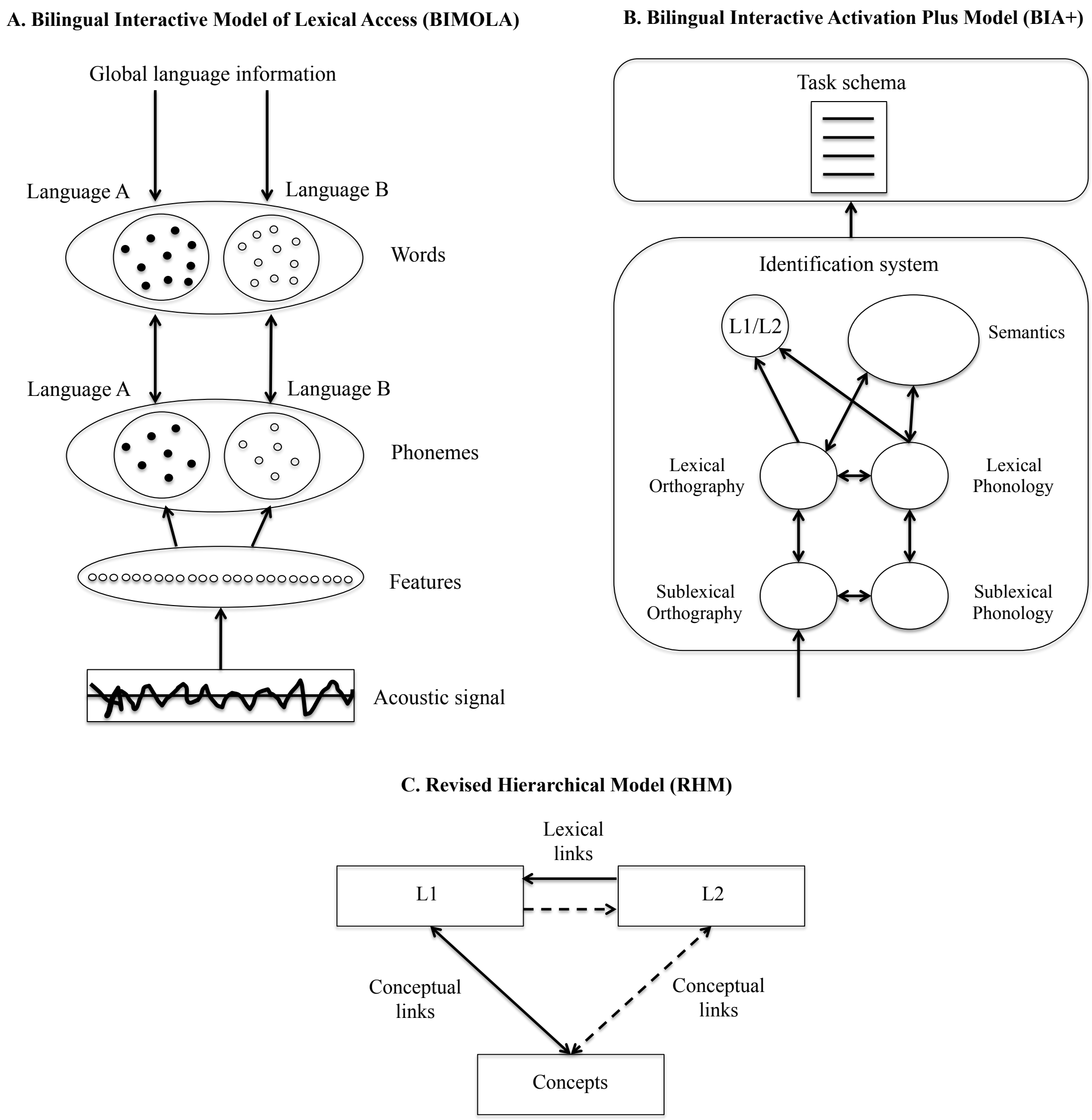
Parents and teachers are often concerned that children with poor literacy skills in their native language will experience similar difficulties when learning a foreign language (Baker, 1996; Genesee, 2015; Hussien, 2014; Palladino, Bellagamba, Ferrari & Cornoldi, 2013; Peer & Reid, 2014; Sparks, 2016; Wight, 2015). However, past research shows that foreign language learning in poor readers/spellers has variable outcomes (Helland & Kaasa, 2005; Ho & Fong, 2005; van der Leij & Morfidi, 2006; von Hagen, Kohnen & Stadie, submitted). For example, in a recent study we found that only 14 out of 32 children with poor literacy skills scored significantly below a group of children with typical literacy skills on several foreign language tasks (von Hagen, Kohnen & Stadie, to be submitted). The remaining 18 poor readers/spellers were just as successful as their classmates with typical literacy skills on all foreign language measures. Nevertheless, the sources of individual variability in foreign language attainment of children with poor literacy skills still remain unknown.

Previous research with children with typical literacy skills shows that native language skills play an important role in explaining individual differences in foreign language learning (Sparks & Ganschow, 1991; Sparks, Ganschow, Javorsky, Pohlman & Patton, 1992a, 1992b; Sparks, Patton, Ganschow, Humbach & Javorsky, 2006; Sparks, Ganschow & Pohlman, 1989). The developmental interdependence hypothesis by Cummins (1978; 1989; 2000) posits that foreign language competence depends to a great extent on native language competence. This hypothesis holds that the performance of a child on a certain native language task (e.g. word reading) predicts performance on an equivalent foreign language task. For example, a child with poor native language auditory discrimination and poor spelling skills might show the same deficits in foreign language auditory discrimination and spelling. From a practical perspective, this information could allow teachers, parents and clinicians to determine which children with poor literacy skills are at greater risk of experiencing foreign language difficulties and what type of difficulties might emerge. Cross-linguistic interdependence, however, has so far not been systematically investigated in children with poor literacy skills.

Cummins (1978; 1989; 2000) suggests that the extent to which native language skills predict parallel foreign language skills, depends on the interaction between common underlying versus language specific resources. Therefore, cross-linguistic interdependence may occur for some linguistic subskills, but not for others. In fact, data from children with typical literacy skills indicates stronger cross-linguistic associations for written than for oral language subskills (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2010). Possibly, written language skills rely to a greater extent on common underlying language resources than oral language skills. Melby-Lervåg and Lervåg (2011) argue that oral language is a more complex and multi-determined domain in which cross-linguistic associations are likely to be harder to detect than in what they describe as “simpler language domains”, such as phonological awareness. The complexity of oral language is captured in theoretical models of language processing that describe distinct processing components within the oral language domain. To the best of our knowledge, a theory-driven analysis investigating cross-linguistic interdependence at different levels of language processing is currently missing. This study addresses this gap by assessing native and parallel foreign language skills of children with poor literacy skills.

Although numerous theoretical models of oral and written bilingual language processing are available in the literature, it is hard to find one single model that encompasses input and output levels and is applicable to oral and written language processing at the same time (Gray & Kiran, 2013). For example, the Bilingual Interactive Model of Lexical Access (BIMOLA - Grosjean, 1988, 1997) aims to account for cross-linguistic speech perception (see Figure 1A).

Figure 1. Examples of models of language processing in two languages.



Note. L1=Native language; L2 = Foreign language. 1A adapted from Thomas & van Heuven (2005); 1B adapted from Dijkstra & van Heuven (2002); 1C adapted from Kroll & Stewart (1994).

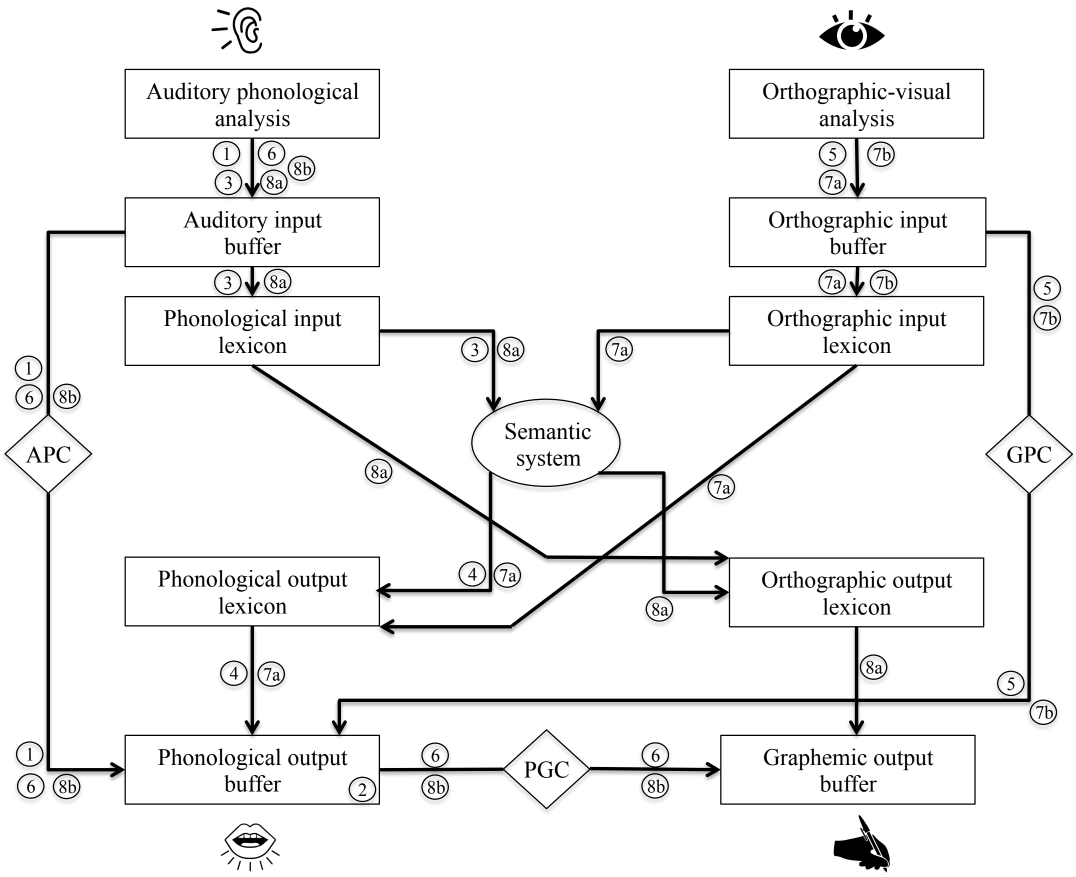
It suggests that the acoustic signal is first processed at a language general feature level. On a second processing level, information on language-specific phonemes is extracted and finally language-specific lexical representations are activated. This model is certainly useful to analyse cross-linguistic interdependence of native and foreign language speech perception, but it does not address other subcomponents of oral and written language processing (e.g. speech production or written language processing).

Several other models also describe native and foreign language spoken word comprehension and production. For example, the Revised Hierarchical Model (RHM - Kroll & Stewart, 1994) proposes that new foreign language words are first linked to corresponding native language words (see Figure 1C). Initially, only an indirect link between the new foreign language word and the semantic representation is established through the mediation of the native language word (e.g. new English foreign word /cʌp/ (cup) ->□□□□□□□□□□ German native word /tasə/ (Tasse) ->semantic representation of cup). As experience with the new foreign language word increases, direct links between the foreign language word and its corresponding semantic representation emerge. Moreover, the model assumes stronger connections from the foreign language word towards the native language word than in the opposite direction (Grainger, Midgley & Holcomb, 2010; Kroll & Stewart, 1994). Supporting evidence comes from studies that have found shorter response times in translations from foreign to native language words than in the opposite direction (Grainger et al., 2010; Kroll & Stewart, 1994). Although the Revised Hierarchical Model provides a useful framework to analyse cross-linguistic interdependence in native and foreign language spoken word comprehension and production, it does not refer to other levels of language processing (e.g. auditory discrimination and speech sound production) that might influence cross-linguistic patterns.

Models focusing on written language processing in two languages also exist in the literature. For example, the Bilingual Interactive Activation Plus model (Dijkstra & van Heuven, 2002) distinguishes between a first processing level of sublexical orthography and phonology and a second level of lexical orthography and phonology (see Figure 1B). Both levels are interconnected. Furthermore, lexical orthography and lexical phonology are linked to a subcomponent of semantic representations and also to "language nodes" that define membership to the native or foreign language. Finally, all components interact with a so-called "task schema", which is responsible for identifying specific requirements of the linguistic task to be completed. Once again, this model has proven useful to interpret data on reading and spelling tasks in two languages but does not address oral language performance.

While the aforementioned models enable theory-driven analyses of native and foreign language processing for distinct subskills, none of them encompasses both input and output levels and is applicable to both oral and written language processing. In order to investigate cross-linguistic interdependence across different subskills, it is imperative to work with a model that fulfils these requirements. One model that is useful in this respect is the Ellis and Young model (1988) of language processing. Although it has not been widely used to explain bilingual language processing, it has been recently proposed by Gray and Kiran (2013) as a theoretical account to investigate language processing deficits in bilingual aphasia. Furthermore, this model (or versions thereof) has a longstanding history of being applied to populations with language deficits, including children with poor literacy skills across different languages (Friedmann, Biran & Dotan, 2013; Friedmann & Coltheart, 2016; Kezilas, Kohnen, McKague & Castles, 2014; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003). Figure 2 depicts an adaptation of Ellis and Young´s (1988) language processing model.

Figure 2. Adaptation of Ellis and Young´s (1988) language processing model.



Note. Adapted from Stadie & Schröder (2009). APC = Auditory phonological conversion; GPC = Grapheme-phoneme conversion; PGC = Phoneme-grapheme conversion; ①Speech sound perception; ② Speech sound production; ③Spoken word comprehension; ④Spoken word production; ⑤ Nonword reading; ⑥Nonword spelling; (7a) Lexical word reading; (7b)Sublexical word reading; (8a) Lexical word spelling; (8b) Sublexical word spelling.

This model details the underlying cognitive mechanisms involved in the following eight subskills that we explored in this study: (1) speech sound perception, (2) speech sound production, (3) spoken word comprehension, (4) spoken word production, (5) nonword reading, (6) nonword spelling, (7) word reading and (8) word spelling. According to this model, speech sound perception involves a first step of auditory-phonological analysis to analyse the acoustic signal. This information is briefly stored in the auditory input buffer, where auditory features are converted into phonemes and then stored in the phonological output buffer (see Figure 2 pathway 1). During speech sound production phonemes are first activated and stored in the phonological output buffer, until specific articulation plans are put in place to pronounce the speech sound (see pathway 2 in Figure 2).

Spoken word comprehension undergoes an initial auditory phonological analysis stage, followed by the involvement of the auditory input buffer to temporarily store information. A matching phonological representation is then activated in the phonological input lexicon and a connection to the corresponding representation in the semantic system is triggered (see pathway 3 in Figure 2). In contrast, for spoken word production, processing begins in the semantic system. Here a conceptual representation is selected for the word the speaker wants to say. The corresponding phonological representation of the word then has to be activated in the phonological output lexicon and passed on to the phonological output buffer. Here, information is stored until a matching articulation plan allows the word to be pronounced (see pathway 4 in Figure 2).

For nonword reading, processing begins with an orthographic-visual analysis, in which letter identity, position and letter-to-word binding take place (Friedmann & Coltheart, 2016). Information is then stored in the orthographic input buffer, before each grapheme is converted into its corresponding phoneme via the "GPC" component. The resulting string of phonemes is then stored in the phonological output buffer, until the necessary articulation plans are in place to pronounce the nonword (see pathway 5 in Figure 2). In contrast, when writing nonwords to dictation, after an initial auditory phonological analysis of the incoming nonword, information is stored in the auditory input buffer. Auditory features are then converted into phonemes and stored in the phonological output buffer. Finally, phonemes are converted into corresponding graphemes and briefly stored in the graphemic output buffer, before matching motor plans are activated to write each letter (see pathway 6 in Figure 2).

Similarly as for nonword reading, word reading also begins with an orthographic-visual analysis and continues to the orthographic input buffer. Depending on the type of word being read (words with regular or irregular grapheme-phoneme correspondences), successful processing of a written word can then either follow a 'lexical' or 'sublexical' processing pathway. Reading words with irregular grapheme-phoneme correspondences requires lexical route processing and involves the activation of word-specific orthographic representations in the orthographic input lexicon. From here, a corresponding phonological representation is either activated directly in the phonological output lexicon or indirectly through the mediation of the semantic system. The selected phonological representation is then stored in the phonological output buffer, until the word is pronounced through the activation of articulation plans (see pathway 7a in Figure 2). In contrast, reading words with regular grapheme-phoneme correspondences can also be achieved via the sublexical pathway (through the conversion of graphemes to phonemes). In this case, the sequence of phonemes resulting from grapheme-phoneme conversion is also stored in the phonological output buffer. Finally, matching articulation plans are activated to read the word aloud (see pathway 7b in Figure 2).

Word spelling can also occur either via lexical or via sublexical pathways. For instance, when spelling words to dictation, processing starts with an auditory-phonological analysis of the spoken word and continues to the auditory input buffer for both words with regular and irregular phoneme-grapheme correspondences. Successful processing of words with irregular phoneme-grapheme correspondences then requires a phonological representation to be activated in the phonological input lexicon. Following this, the corresponding orthographic representation in the orthographic output lexicon needs to be activated either directly or indirectly (through the semantic system). Finally, this orthographic representation is temporarily held in the graphemic output buffer until matching motor plans are put in action to write the word (see pathway 8a in Figure 2). In contrast, spelling words with regular phoneme-grapheme correspondences can either follow the same lexical processing path as for words with irregular phoneme-grapheme correspondence or a sublexical processing path can be activated. In the case of sublexical processing, the sequence of phonemes stored in the phonological input buffer is passed on to the phonological output buffer. Here, information is briefly stored to allow for phoneme-grapheme conversion. The resulting sequence of graphemes is then again held in the graphemic output buffer until corresponding motor plans enable the word to be written (see pathway 8b in Figure 2).

Processing pathways for word reading and spelling can vary as a function of the characteristics of the writing system of each language (Geva & Siegel, 2000; Katz & Frost, 1992; Yap & Rickard Liow, 2016). While more irregular orthographies, such as English, rely predominantly on lexical processing, in more regular orthographies, such as German, successful reading and spelling for most words can occur either through lexical or sublexical processing (Frith, Wimmer & Landerl, 1998; Landerl, 2017; Landerl, Wimmer & Frith, 1996).

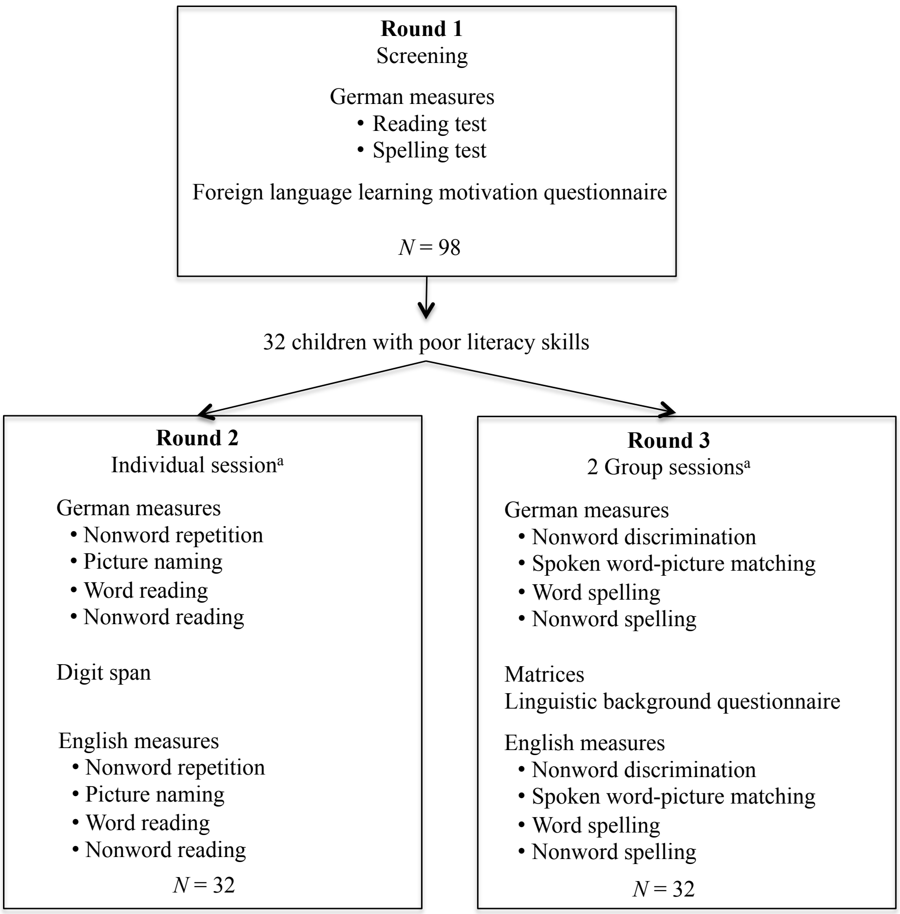
The primary aim of this study was to investigate cross-linguistic interdependence in children with poor literacy skills, based on Ellis and Young's (1988) model of language processing. More specifically, we intended to assess if native language subskills contribute to explaining individual differences in equivalent foreign language subskills in poor readers/spellers. Therefore, we collected data on eight native and foreign language subskills in a sample of 32 German native speaking children with poor literacy skills, who were learning English as a foreign language. In addition to native language skills, past research with children with typical literacy skills also indicates that linguistic background (monolingual vs. bilingual), intellectual ability, short term and working memory and foreign language learning motivation play a role in explaining individual differences in foreign language attainment (Cenoz, 2013; Csizér & Magid, 2014; Dörnyei & Ryan, 2015; Gardner, 1985; Li, 2016; Linck, Osthus, Koeth & Bunting, 2014). Therefore, as a secondary aim, we explored the role of these broader cognitive measures as potential sources of individual variability in foreign language attainment of children with poor literacy skills.

Method

Schools and Participants

Ethics approval was obtained from the University of Potsdam and the education ministry of Berlin. This study was part of a broader study described in von Hagen, Kohnen and Stadie (to be submitted). While the complete data collection process included the recruitment of children with typical and poor literacy skills, in the present study we only focused on the group of children with poor literacy skills (see Figure 3 for data collection process).

Figure 3. Flowchart of the data collection process.



Note. [a]We administered all tests in one language together in order to facilitate the language modus activated by the child (Grosjean, 1998). However, we systematically varied the language blocks and the order of the tasks within each language block across participants.

Participating schools. We invited 50 public primary schools in Berlin, Germany, to participate in the project. Non-mainstream schools with specific pedagogies were excluded (e.g., Montessori: Montessori (2013), Waldorf: Steiner (1997)). Four schools agreed to participate in our study.

While some schools reported that their students were already exposed to English as a foreign language in grades 1 and 2 in an unsystematic playful way (e.g., by singing English Christmas songs), all schools began systematic English instruction in grade 3. From this time onwards, the amount of English instruction is regulated by the primary school authority of Berlin. Therefore, all students were exposed to the same amount of English instruction starting with two lessons of approximately 40-45 minutes per week in grade 3. English instruction increased by one lesson a week resulting in five lessons a week in grade 6. The students did not receive any other foreign language instruction within the school context.

Following Müller (2010), we used the percentage of students exempted from paying an extra contribution for learning materials as an approximation of the school´s socio-economic status (SES). Exemptions are granted to students of families that receive financial support from the government due to a low family income. Only 10 primary schools of Berlin have an exemption rate of 90% or more of their students, a quarter of primary schools have an exemption rate of 53% or more and half have an exemption rate of 38% or more (Müller, 2010). In the current study, two of the participating schools belonged to neighbourhoods in which up to 80% of the students are exempted from this fee, while the other two were located in areas with an exemption percentage lower than 40%.

Participating children. We contacted the parents of 224 students, of which 101 parents agreed to their child´s participation. Of these 101 children, 98 children signed a written assent form agreeing to be tested. Both parents and children knew that the study was investigating foreign language learning in children with poor and typical literacy skills. The students attended 10 different grade 6 classrooms and received English as a foreign language instruction by nine different teachers. Based on the information provided in a parent questionnaire on the children's developmental history, we had planned to exclude children with uncorrected hearing or sight impairments and bilingual children with English as one of the native languages. However, none of the children who agreed to participate in the study met these exclusion criteria.

The 98 children completed a reading and spelling screening test. Thirty-two children scored in the lowest 20% of the norm comparison group either on the reading or the spelling screening test or on both and were identified as children with poor literacy skills. This cut-off criterion was chosen based on the definition of poor readers in the manual of the ELFE 1-6, the standardized reading test that we used (Lenhard & Schneider, 2006). The final participant group consisted of 15 girls and 17 boys. While 22 of the children spoke only German as a native language, the parents of 10 students reported that their children also spoke another language at home (i.e. Albanian, Arabic, Kurdish, Persian, Polish, Spanish and Turkish). The children's mean age was $M = 11.86$ years; $SD = 0.44$ and their performance on the literacy screening measures was $M = 29.50$; $SD = 5.00$; $Mdn = 30.50$ in raw scores and $M = 21.78$; $SD = 15.40$; $Mdn = 21.18$ in percentiles for the reading screening test and $M = 29.50$; $SD = 8.68$; $Mdn = 28.50$ in raw scores and $M = 22.58$; $SD = 12.00$; $Mdn = 24.39$ in percentiles for the spelling screening test.

Procedure. We conducted three "testing rounds" (see Figure 3). Ninety-eight children participated in Round 1, during which we assessed native language reading and spelling skills to identify children with poor literacy skills. In addition, in Round 1 children completed a questionnaire on their foreign language learning motivation. We identified 32 children with poor literacy skills who completed testing Rounds 2 and 3. In Round 2 we collected data on several native and foreign language tasks and a digit span task as a measure of verbal short and working memory. Round 3 focused on additional native and foreign language tasks, as well as the Matrices subtest, as a measure of nonverbal reasoning. Lastly, children were asked to complete a questionnaire on their linguistic background in Round 3.

For Round 1, we administered the tests in group-settings (by classroom), while Round 2 consisted of one individual session. Finally, Round 3 comprised two group sessions (max. of 12 children). Each test session lasted approximately 45 minutes. The instructions were given in the language the children were being tested in and the first author or a research assistant specifically trained for this project completed the assessment. Testing occurred in the second half of the school year, during regular school time, in a quiet room assigned by the school. Both testers were native German speakers with proficient knowledge of English.

Materials

Native language literacy screening measures.

Reading test. To identify poor German native language literacy skills we used the word reading subtest from the *ELFE 1-6 Ein Lese-verständnistest für Erst- bis Sechstklässler* (Lenhard & Schneider, 2006). Participants had to choose the correct written target word between four phonemically and graphemically similar distractor options that matched a picture. For example, they saw the picture of a window and had to mark *Fenster* (*window* in English) as the correct option. Distractor items were *Felsen* (*rock* in English), *Fehler* (*mistake* in English) and *Fremder* (*stranger* in English). Following the standard procedure, we gave children two minutes to complete as many items as possible out of a total of 72. The internal consistency of the different subtests of this test varies between α = .92 and α = .97 and the test-retest reliability measured after 14 days was r_{tt} = .91, with a validity of r = .71 (Lenhard & Schneider, 2006). Poor readers were defined as children who performed at or below the 20th percentile on this measure.

Spelling test. We used the *Hamburger Schreibprobe (HSP 5-10)* (May, 2012) to identify poor German native language spelling skills. More specifically, we focused on the number of correctly spelled words score of this test. Children were asked to complete a spelling to dictation task of 14 words, five sentences and a sentence completion task. Internal consistency for this measure is reported as between $r = .92$ and $r = .99$, and test-retest reliability is $r_{tt} = .52$ to .93, depending on the retest timing; validity is reported as $r = .87$ (May, 2012). Students who performed at or lower than the 20th percentile on the number of correctly written words score were selected as poor spellers. Typical spellers scored above the 20th percentile on this test.

Measures of potential sources of individual differences in foreign language attainment of children with poor literacy skills.

Native language skills. We used bespoke and standardized tasks to assess participants on the following native language measures: speech sound perception and speech sound production, spoken word comprehension and production, nonword reading and spelling, and word reading and spelling. A description of all measures is provided in Table 1.

Table 1
Description of native language tasks

Task	N of items	Description
Nonword discrimination	20	In this experimenter-designed task, children listened to a minimal pair of nonwords and were asked to mark on a response sheet if the pair was identical or not. The stimuli were recorded by a native speaker of German and contained only one German specific phoneme, which does not occur in English and filler phonemes, common to both languages. Half of the pairs were identical and half differed with respect to one phoneme. We selected the phoneme contrasts based on previous studies on foreign language sound perception (Bohn, Best, Avesani, & Vayra, 2011; Bohn & Flege, 1992; König & Gast; 2012; Trehub, 1976). See the supplemental materials for more details on the phonological contrasts that were used and for the list of the items. Internal consistency was measured by Cronbach's α and was α = .25.
Nonword repetition	24	In this experimenter-designed task, children were asked to repeat a nonword that had been recorded by a native German speaker. Each nonword contained one German specific phoneme that does not exist in English (König & Gast; 2012). These target phonemes occurred once in a mono-, a bi- and a tri-syllabic nonword respectively. The filler phonemes were common to both languages. The mean number of phonemes in the items was 5.25 with a range of 3-8. See the supplemental materials for the list of items. Children's responses were recorded and scored by two native German speakers. They judged if the language specific phoneme was produced in a native-like way or not. The double scoring of 45% of the recordings revealed an average inter-rater-reliability of κ = .87. The recording of one participant was deleted by error before the scoring procedure. Therefore we only had data of this measure on n = 31 poor readers/spellers.
Spoken word-picture matching	40	We used the short version of the *WWT 6-10* (Glück, 2011), which assesses the same items (20 nouns, 10 verbs, 10 adjectives) in the receptive and expressive modalities. The receptive task was a spoken word-picture matching task where children heard a word recorded by a German native speaker. Children were asked to select the picture (out of four choices with phonological, semantic or unrelated distractors) which best represented the word they had heard. We adapted the task to better suit group administration by projecting the pictures onto a screen. For the picture naming task (expressive modality), children saw a picture and their response was elicited through a question (e.g. What is this?, What is the opposite of…?, What is she doing?). The internal consistency value reported for this short version of the published test is α = .84 (Glück, 2011).
Picture naming	40	
Word reading	max. 156	We selected the word and nonword reading subtests from the *SLRT II* (Moll & Landerl, 2010). First, children were asked to read a list of words as quickly and accurately as they could for one minute. The same procedure was followed with a list of nonwords derived from the word list. The total score of each task corresponds to the number of correctly read items. The reported test-retest reliability values are r = .90 - .98 (Moll & Landerl, 2010).
Nonword reading	max. 156	
Word spelling	20	We adapted the word and nonword spelling subtests from the adult language assessment battery *LEMO 2.0* (Stadie, Cholewa & De Bleser, 2013). We chose the 20 monosyllabic irregular words from the word spelling subtest and the 20 monosyllabic regular nonwords derived from the word items in the nonword spelling subtest. Children heard the words from a recording spoken by a native speaker and were asked to write their response on a piece of paper. The internal consistency value based on the performance of the participants in this study corresponds to α = .73 for the word spelling task and to α = .80 for the nonword spelling task.
Nonword spelling	20	

Linguistic background.

Parent questionnaire. We asked parents to complete a questionnaire with a multiple choice and brief answer format on their child's developmental history, their linguistic background and the presence of any learning difficulties. This information was used to categorize children as monolingual, if parents reported that German was the only language spoken at home, or as multilingual, if children were exposed to other languages at home.

Student questionnaire. This instrument was added to the test battery after testing had started, because some children reported additional knowledge of languages, which had not been mentioned by their parents. In a short questionnaire, we asked children about their linguistic background, their self-perception of their competencies in the languages they speak, and when they had started learning English. This information was used to complement the data provided by parents. Both questionnaires are attached in the supplemental materials.

Foreign language learning motivation. To assess children´s motivation to learn English we adapted the *Attitude/Motivation Test Battery (AMTB)* by Gardner (1985). The following items were selected from the original questionnaire and translated to German: six items from the dimension *Attitudes towards learning English* (French in the original); two items from the *Integrative Orientation* construct; two items from the *Instrumental Orientation* dimension; four items measuring *English class anxiety* (French in the original); six items of the *Motivational Intensity* dimension. The resulting 20 items were read aloud to the children. For the first 14 items they were asked to mark how much they agreed with each statement on a Likert scale with seven options ranging from "not at all" to "totally agree". Following the original version of the AMTB, we assigned a score ranging from minus three to three for each item. For the last six items we asked participants to mark one out of three multiple-choice options, which received a score from one to three. The total score ranged from -36 to 60 with a higher score corresponding to higher foreign language learning motivation. Internal consistency as measured by Cronbach's α was .61 in our sample of 32 poor readers/spellers. The full questionnaire is attached in the supplemental materials.

Nonverbal reasoning. We used the subtest *Matrices* from the German Version of the *Wechsler Intellectual abilities Scale for Children - Fourth Edition (WISC IV)* (Petermann & Petermann, 2014). In order to adapt this task to be administered to a group, we projected the items onto a screen. Children saw an incomplete puzzle with five options for the missing piece and were asked to mark the correct option on the corresponding numbered box on a response sheet. To take into account the starting point indicated in the test manual for our participant group´s age range, we presented items number seven to 35 of the original task. Although children were asked to complete all items, the standard discontinuing rule indicated by the test manual was applied when scoring. Despite having violated the individual administration conditions under which the norms were constructed, we still decided to use the test norms as a reference instead of the raw scores, as they take into account participant age. Split-half reliability is reported as $r = .89$ to $r = .91$ for children from 10 to 12 years of age according to the test manual (Petermann & Petermann, 2014).

Verbal short and working memory. This measure was assessed through the subtest *Forward and Backward Digit Span* from the German version of the *Wechsler Intellectual abilities Scale for Children - Fourth Edition (WISC IV)* (Petermann & Petermann, 2014). In the first part of this task, children were asked to repeat a sequence of numbers that they were orally presented with. In the second part, they had to repeat a sequence of numbers in the reverse order. The length of the number chain increased as the task progressed. Children heard the items from a recording on the computer to ensure that the presentation was the same for all participants. Split-half reliability for children from 10 to 12 years of age was reported to be between $r = .72$ to $r = .82$ and $r = .71$ to $r = .81$ for the forward and backward digit span task, respectively (Petermann & Petermann, 2014).

English as a foreign language outcome measures. To assess English foreign language attainment we designed parallel tasks to the native language tasks described in Table 1. All measures are described in Table 2.

Table 2

Description of English as a foreign language tasks

Task	N of items	Description
		All foreign language tasks were specifically designed considering the contrasts between German and English and taking into account the foreign language textbook vocabulary of the participants.
Nonword discrimination	26	Children listened to a minimal pair of nonwords and were asked to mark on a response sheet if the pair sounded identical or not. Twenty items contained only one phoneme that only exists in English and does not occur in German and filler phonemes, common to both languages. Six additional items were added that contrasted an English specific phoneme with a German specific phoneme. Overall, half of the pairs were identical and half differed with respect to one phoneme. We selected the phoneme contrasts based on previous studies on foreign language sound discrimination (Bohn, Best, Avesani, & Vayra, 2011; Bohn & Flege, 1992; König & Gast; 2012; Trehub, 1976). See the supplemental materials for the list of the items. The stimuli were recorded by a native German-English bilingual speaker with English as a dominant language. The internal consistency value measured through Cronbach's α corresponds to α = .29.
Nonword repetition	24	Each child was asked to repeat a nonword that had been recorded by a native German-English bilingual speaker with English as a dominant language. Each nonword contained one English specific phoneme that does not exist in German (König & Gast; 2012). These target phonemes occurred once in a mono-, once in a bi- and once in a tri-syllabic nonword. The filler phonemes were common to both languages. The mean number of phonemes across all items was 5.25 with a range of 3-8. See the supplemental materials for a list of items. Two native English speakers judged the native likeliness of the responses focusing on the target phoneme contrast. A double scoring of 22% of the recordings revealed an inter-rater reliability of κ = .77.
Spoken word-picture matching	48	48 items were selected from the FL textbooks used by the different schools from grade 3 to grade 6 (8 different books). We only used items present in all textbooks and marked as core vocabulary. The same items were used for the receptive and expressive tasks. We controlled for the following psycholinguistic variables, which have been shown to influence second language vocabulary learning: grammatical class, time point of first exposure during FL instruction, similarity to first language (cognate status), presence of consonant clusters and syllable length (see the supplemental materials for details). For the receptive task, children heard a word recorded by a German-English bilingual speaker with English as a dominant language and were asked to mark the correct option between four pictures on a response sheet. For the expressive task children saw two pictures and heard a sentence for the first picture. Their response to the second picture was elicited by an incomplete sentence (e.g. This is slow. This is……). The internal consistency value as measured by Cronbach's α corresponds to α = .81 for the receptive task and to α = .84 for the expressive task.
Picture naming	48	
Nonword reading	24	We designed a list of 24 monosyllabic nonwords including twelve grapheme-phoneme correspondences (GPC) that differ between German and English. Each contrast occurred in two nonwords. Three types of GPC contrasts between both languages were distinguished: (a) new foreign language phoneme - common native and foreign language grapheme, (b) new foreign language grapheme -common native and foreign language phoneme and (c) common native and foreign language phoneme and grapheme, but new GPC (see the supplemental materials). We asked children to read the items aloud. Children's responses were audio-recorded and transcribed after the testing session. The total score corresponds to the correctly read target GPC. The internal consistency value corresponds to α = .75 measured by Cronbach's α.

Task	N of items	Description
Nonword spelling	24	We used the same nonword list as in the nonword reading task. Items were presented by a recording from a native speaker through a loudspeaker and children were asked to write the nonwords on a response sheet. Only the target phoneme-grapheme correspondence (PGC) in each item was scored as correct or incorrect according to the following criteria, which have been previously used by Kohnen, Colenbrander, Krajenbrink & Nickels (2015): only PGC with a type frequency of 20 and a token frequency of at least 20000 out of all words in the CELEX database (Baayen et al., 1993) were scored as correct. The internal consistency value measured by Cronbach's α corresponds to α = .56.
Word reading	24	We selected irregular words from the FL textbook vocabulary. Words were considered irregular if their dictionary pronunciation based on the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) was different from a rule-based pronunciation based on the DRC rules (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). Furthermore, the following psycholinguistic variables were controlled: time point of exposure during foreign language instruction, cognate status and presence of consonant clusters (see the supplemental materials for the list of items). Children were asked to read the words aloud. The internal consistency value measured as Cronbach's α corresponds to α = .84.
Word spelling	24	The same word list as in the foreign language word reading task was used. Children heard the items that were recorded by a native speaker through a loudspeaker and were asked to write them on a response sheet. Word spellings were either scored as correct or incorrect. The internal consistency value measured by Cronbach's α corresponds to α = .84.

Results

First, we present the descriptive statistics for all measures (see Table 3).

Table 3

Descriptive statistics for all experimental measures

	M (SD)	*Mdn*	*Min*	*Max*
German native language measures[a]				
Nonword discrimination (*n* = 20)	19.03 (1.10)	19.00	15.00	20.00
Nonword repetition[b] (*n* = 24)	16.68 (3.16)	17.00	10.00	23.00
Spoken word-picture matching (*n* = 40)	36.47 (3.00)	37.00	24.00	40.00
Picture naming (*n* = 40)	27.16 (7.37)	29.00	12.00	38.00
Nonword reading (*n* = max. 156)	35.75 (11.04)	35.50	6.00	62.00
Nonword spelling (*n* = 20)	12.28 (3.21)	12.00	4.00	19.00
Word reading (*n* = max. 156)	66.91 (21.96)	65.50	29.00	110.00
Word spelling (*n* = 20)	14.69 (3.01)	15.50	8.00	19.00
Foreign language learning motivation questionnaire[c] (*n* = 20)	33.44 (6.09)	32.50	22.00	46.00
Matrices[d] (*n* = 35)	10.78 (2.30)	11.00	5.00	14.00
Forward digit span[d] (*n* = 16)	7.25 (1.37)	7.00	5.00	10.00
Backward digit span[d] (*n* = 16)	6.63 (1.27)	6.00	4.00	10.00
English foreign language measures[a]				
Nonword discrimination (*n* = 26)	19.47 (1.79)	20.00	14.00	23.00
Nonword repetition (*n* = 24)	20.41 (1.73)	20.00	16.00	23.00
Spoken word-picture matching (*n* = 48)	42.19 (4.37)	44.00	28.00	48.00
Picture naming (*n* = 48)	34.03 (6.27)	34.50	10.00	42.00
Nonword reading (*n* = 24)	13.22 (3.68)	14.00	6.00	17.00
Nonword spelling (*n* = 24)	12.03 (2.92)	12.00	6.00	23.00
Word reading (*n* = 24)	15.41 (4.83)	15.00	1.00	23.00
Word spelling (*n* = 24)	12.13 (4.84)	12.00	1.00	22.00

Note. The number of items for each measure are presented in parentheses. [a]Expressed in raw scores. [b]The data for 32 children with poor literacy skills was available for all the tests with the exception of the German nonword repetition task. In this task the recording of one child was deleted by error and so the results are based on *n* = 31. [c]Each item received a score from -3 to 3 based on the participant's response on a Likert scale. Total scores could therefore range from -36 to 60. [d]Expressed in scaled scores based on the WISC IV test manual (Petermann & Petermann, 2014).

In order to investigate if native language subskills and broader cognitive measures contribute to explaining individual differences in foreign language attainment in poor readers/spellers, we used linear mixed-effects modelling in R (Baayen, 2008), as implemented in lme4 package (Bates, Maechler, Bolker, & Walker, 2014). We entered participant and test item intercepts as random factors into the model. The fixed factors were the following experimental measures: German native language skills (i.e. speech sound perception and speech sound production, spoken word comprehension and production, word and nonword reading and spelling), linguistic background, foreign language learning motivation, nonverbal reasoning, verbal short term memory, verbal working memory. These variables were rescaled to $M = 0$ and $SD = 1$ to obtain comparable regression coefficients.

Variance inflation factors for each fixed effect indicated that multi-collinearity was present for the following three German native language fixed effects (VIF > 3): spoken word production (VIF = 3.64), word reading (VIF = 4.13) and nonword reading (VIF = 3.40). As we were interested in investigating the contribution of these native language measures to explain individual differences in equivalent foreign language measures in poor readers/spellers, we decided to include these fixed effects in our analyses despite the potential multi-collinearities. However, we will return to the implications of this decision in the Discussion.

The model was run separately for each of the English foreign language outcome measures (i.e. speech sound perception, speech sound production, spoken word comprehension, spoken word production, nonword reading, nonword spelling, word reading and word spelling). In all cases, the dependent variable was item-level accuracy, so we fit logistic regression models.

We started the analysis with a full model containing the 13 fixed effect variables. Next, we entered a reduced model by removing all the non-significant variables from the full model. In order to select this reduced model as the optimal model, two criteria had to be met:

(a) the reduced model had to be just as efficient as the full model in explaining the variance in the foreign language outcome measure, and (b) the reduced model had to explain more variance than a null model without any fixed effects. To test these conditions we compared the reduced model to (a) the full model and (b) the null model using a chi-square test of the variance explained.

If the reduced model fulfilled both criteria, we tried to simplify the model by removing further fixed effects in order of their contribution in explaining the outcome variable. We only accepted a simpler model as the optimal model if it met the aforementioned conditions.

All of the analyses were first completed with a database of 31 children, because the data for native language speech sound production of one child was deleted by error. However, as results showed no significant contribution of native language speech sound production to explain the observed variance in any of the eight foreign language outcome measures, we re-ran the analysis with the original database of 32 children and excluded native language speech sound production from the model. Following, we report the results for each of the dependent variables in turn.

With respect to foreign language speech sound perception, the outcome variable was correct or incorrect detection of whether two nonwords were the same or different (see Table 2 in Methods). The analysis was based on 832 observations in total for 32 children and 26 items. The results show that a higher score on the native language measure was associated with a higher score on the foreign language measure (see Table 4). This optimal model differed significantly from the null model, $\chi^2(1) = 8.084$; $p = .004$, but not from the full model, $\chi^2(11) = 7.06$; $p = .794$.

Table 4

Optimal logistics mixed effects model for foreign language speech sound perception

Fixed effects	FL auditory discrimination of minimal pairs			
	Estimate	*SE*	*z*	*p*
Intercept	-1.795	1.695	-1.059	.289
NL speech sound perception	0.223	0.087	2.567*	.010*

Note. FL = Foreign language*;* *SE* = Standard error*;* NL = Native language

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

For foreign language speech sound production the outcome variable was native-like or non native-like repetition of a target foreign language phoneme embedded in a nonword. English natives speakers judged native-like repetitions as correct and non native-like repetitions as incorrect for each item (see Table 2 in Methods). A total of 768 observations from 32 children and 24 items were included in the analysis. No optimal model was found, as none of the 12 fixed effects showed a significant contribution in explaining the observed variance in foreign language speech sound production scores.

Regarding foreign language spoken word comprehension the outcome variable was correct or incorrect identification of a spoken word-picture match (see Table 2 in Methods). In total 1536 observations from 32 children and 48 items were analysed. None of the 12 fixed effects showed a significant contribution in explaining the observed variance in foreign language spoken word comprehension. Thus, no optimal model was found.

For foreign language spoken word production the outcome variable was the correct or incorrect naming of a picture (see Table 2 in Methods) and 1536 observations from 32 children with 48 items were analysed. Higher scores in this task were significantly associated with higher native language word reading and spoken word production skills and foreign language learning motivation (see Table 5). The optimal model differed significantly from the null model, $\chi^2(3) = 27.149$; $p < .001$, but not from the full model, $\chi^2(9) = 8.267$; $p = .507$.

Table 5

Optimal logistics mixed effects model for foreign language spoken word production

Fixed effects	FL spoken word production			
	Estimate	*SE*	*z*	*p*
Intercept	1.791	0.408	4.386	<.001***
NL word reading	0.651	0.146	4.445	<.001***
NL spoken word production	0.533	0.149	3.566	<.001***
FL learning motivation	0.468	0.148	3.160	.001**

Note. FL = Foreign language; *SE* = Standard error; NL = Native language
* $p < .05$; ** $p < .01$; *** $p < .001$

In the case of foreign language nonword reading, the outcome variable was correct and incorrect reading of a nonword (see Table 2 in Methods) and the analysis was based on 768 observations from 32 children and 24 items. A higher score on this measure was significantly associated to higher native language word reading skills (see Table 6). The optimal model differed significantly from the null model, $\chi^2(1) = 15.031$; p < .001, but not from the full model, $\chi^2(11) = 10.761$; $p = .463$.

Table 6

Optimal logistics mixed effects model for foreign language nonword reading task

Fixed effects	FL nonword reading			
	Estimate	*SE*	*z*	*p*
Intercept	0.259	0.268	0.967	.334
NL word reading	0.526	0.123	4.274	<.001***

Note. FL = Foreign language; *SE* = Standard error; NL = Native language; * $p < .05$; ** $p < .01$; *** $p < .001$

Regarding foreign language nonword spelling the outcome variable was correct or incorrect spelling of the item (see Table 2 in Methods) and 768 observations from 32 children and 24 items were included in the analysis. A better performance on this measure was significantly associated with higher native language nonword spelling skills (see Table 7). The optimal model differed significantly from the null model, $\chi^2(1) = 7.045$; $p = .008$, but not from the full model, $\chi^2(11) = 16.34$; $p = .129$.

Table 7

Optimal logistics mixed effects model for foreign language nonword spelling task

Fixed effects	FL nonword spelling			
	Estimate	*SE*	*z*	*p*
Intercept	-0.002	0.344	-0.008	.993
NL nonword spelling	0.324	0.116	2.797	.005**

Note. FL = Foreign language; *SE* = Standard error; NL = Native language; * $p < .05$; ** $p < .01$; *** $p < .001$

The optimal model for foreign language word reading skills was based on 744 observations from 32 children and 24 items on the outcome variable correct and incorrect reading of a word (see Table 2 in Methods). A better performance on this measure was significantly associated with higher native language word reading, spoken word production and word spelling skills (see Table 8). This optimal model differed significantly from the null model, $\chi^2(3) = 26.424$; $p < .001$, but not from the full model, $\chi^2(9) = 4.730$; $p = .857$.

Table 8

Optimal logistics mixed effects model for foreign language word reading task

Fixed effects	FL word reading			
	Estimate	*SE*	*z*	*p*
Intercept	0.850	0.331	2.570	.010*
NL word reading	0.753	0.172	4.376	<.001***
NL spoken word production	0.538	0.167	3.206	.001**
NL word spelling	0.460	0.166	2.773	.005**

Note. FL = Foreign language; *SE* = Standard error; NL = Native language * $p < .05$; ** $p < .01$; *** $p < .001$

Lastly, with respect to foreign language word spelling, correct or incorrect spelling of a word was coded as the outcome variable (see Table 2 in Methods). The analysis was based on 768 observations on 32 children and 24 items. We first obtained a model including native word reading, word spelling, nonword spelling and foreign language learning motivation that significantly differed from the null model, $\chi^2(4) = 42.567$; $p < .001$, but not from the full model, $\chi^2(8) = 9.475$; $p = .304$. In order to avoid over-specifying the variation on the data of only 32 participants, we decided to limit the model to three fixed factors (see Paradis & Jia, 2017, for a similar procedure). Therefore, we compared all possible three-way combinations of the four fixed effect variables that we had found to be significant in the model mentioned above. The final optimal model was chosen based on the lowest AIC value. This model indicated that higher foreign language word spelling skills were associated with higher native language word reading, spelling and nonword spelling skills (see Table 9). This final model differed significantly from the null model, $\chi^2(3) = 35.832$; $p < .001$, but not from the full model, $\chi^2(9) = 16.211$; $p = .063$.

Table 9

Optimal logistics mixed effects model for foreign language word spelling task

Fixed effects	FL word spelling			
	Estimate	*SE*	*z*	*p*
Intercept	0.014	0.278	0.959	.959
NL word reading	0.545	0.139	3.920	<.001***
NL word spelling	0.442	0.131	3.356	<.001***
NL nonword spelling	0.350	0.147	2.379	.017*

Note. FL = Foreign language; *SE* = Standard error; NL = Native language; * $p < .05$; ** $p < .01$; *** $p < .001$

To summarize, we found significant contributions of native language speech sound perception, spoken word production, nonword spelling, word reading and word spelling skills to explain individual differences in equivalent foreign language measures in children with poor literacy skills. In contrast, our results revealed no significant impact of native language speech sound production, spoken word comprehension and nonword reading on the same foreign language subskills. Furthermore, only foreign language learning motivation, but not linguistic background, nonverbal reasoning, verbal short term and working memory accounted for the observed variance in poor reader/spellers' foreign language performance.

Discussion

The main aim of this study was to investigate if native language subskills contribute to explaining individual differences in the performance of children with poor native language literacy skills on foreign language subskills. To achieve this goal, we performed a model-based analysis of cross-linguistic interdependence between native and foreign language skills, integrating current models of bilingual processing into Ellis and Young's model (1988) of language processing. As a secondary aim, we also explored the role of broader cognitive measures as potential sources of individual variability in foreign language attainment of poor readers/spellers.

Our results revealed a significant contribution of native language speech sound perception, spoken word production, nonword spelling, word reading and word spelling skills to explain individual differences in corresponding foreign language measures in poor readers/spellers. In contrast, native language speech sound production, spoken word comprehension and nonword reading did not account for the observed variability. These findings are consistent with previous research with children with typical literacy skills in revealing cross-linguistic interdependence for some subskills, but not all (Melby-Lervåg & Lervåg, 2011; Proctor et al., 2010; Verhoeven, 1994). What are the underlying cognitive mechanisms that explain these contrasting results? To answer this question, we discuss our findings within models of language processing.

First, we found native language speech sound perception to play a significant role in explaining individual differences in the same foreign language subskill. According to Ellis and Young's model (1988), speech sound perception involves acoustic analyses and phonemic identification (see pathway ① in Figure 2 in the Introduction). Bilingual processing models, such as the Bilingual Interactive Model of Lexical Access (BIMOLA - Grosjean, 1988, 1997), assume a more prominent involvement of common underlying resources during acoustic analysis, as opposed to a larger reliance on language-specific resources during phonemic identification (see Figure 1A in Introduction - Best, 1995; Best & Tyler, 2007; Flege, 1995). Our findings could therefore be reflecting a larger influence of common underlying resources at the acoustic analysis level, as opposed to the involvement of language-specific resources during phoneme identification. This would mean that individual differences, for example, in the ability to rely on acoustic cues to differentiate fricative and plosive speech sounds could impact the discrimination of language-specific phonemes of these sound categories to a similar extent in native and foreign language speech sound perception (e.g. German specific contrast between fricative /x/ and plosive /k/ in minimal pair /pʊ**x**/ vs. /pʊ**k**/ and equivalent English specific example /**θ**iːp/ vs. /**t**iːp/ - see supplemental materials for details).

However, two issues limit the interpretation of our results. First, we cannot be sure that our findings reflect cross-linguistic interdependence at the acoustic processing stage (as assumed by BIMOLA), as our nonword discrimination task involved not only acoustic, but also phonemic processing. Future studies testing more specifically both processing stages of speech sound perception (acoustic and phonemic discrimination) are needed to shed light on this issue. Second, performance on the native language nonword discrimination had a restricted range because the task produces near-ceiling performance and therefore limits the interpretation of our results. One of the difficulties in speech sound perception in middle to late primary school children (aged 11-12 years) is that they are likely to have mastered this subskill in their native language. Therefore, ceiling performance is probable (even if some of the poor readers/spellers might have experienced difficulties in native language speech sound perception at an earlier age). We chose this age group because we were interested in studying a population with considerable exposure to the foreign language. However, future examinations of auditory discrimination may need to test younger populations for whom native language skills of this kind are still developing.

To the best of our knowledge, this is the first study exploring cross-linguistic interdependence between native and foreign language speech sound perception in children with poor literacy skills. While Soroli, Szenkovits and Ramus (2010) completed similar native and foreign language nonword discrimination tasks with adults with dyslexia, they did not aim to investigate cross-linguistic interdependence and therefore do not report relevant data in this respect. Instead, they report that adults with dyslexia were just as successful as their peers with typical literacy skills in discriminating native and foreign language phonemes.

Turning now to speech sound production, we found no evidence of cross-linguistic interdependence between native and foreign language speech sound production skills in children with poor literacy skills. Ellis and Young's model (1988) suggests that two main processing steps are involved in speech sound production: phonemic activation and articulation (see Figure 2 in Introduction). In a similar way as for phonemic information is assumed to rely predominantly on language-specific resources (Best, 1995; Best & Tyler, 2007; Flege, 1995), Simmonds, Wise and Leech (2011) suggest that articulation requires language-specific oral motor movements. Motor movements for native language sound production are highly over-learned and automatic, while articulation plans would be less developed for non-native phonemes. Therefore, both processing steps involved in native and foreign language speech sound production seem to rely to a greater extent on language-specific than common underlying resources – at least where phonemes differ across languages, (and this was the focus in the nonword repetition tasks used in this study). This means that for example, individual differences in the ability to produce the German-specific sound /ʁ/ in /**ʁ**ɪm/ would have no impact on the production of an English-specific sound, such as /ɹ/ in /**ɹ**ɪʃ/ (see supplemental materials for details) in our study. This is consistent with our results. While previous studies have investigated cross-linguistic interdependence with respect to the production of sounds that are common to both languages in bilingual children with language disorders (e.g. Holm & Dodd, 2006), we believe this to be the first study addressing this issue in poor readers/spellers focusing on the production of language-specific sounds.

Another finding of this study is that individual differences in native language spoken word production, but not comprehension contributed to explaining variation in the corresponding foreign language skills. According to Ellis and Young's (1988) model, the phonological input and output lexicon and the semantic system are key components in spoken word comprehension and production (see Figure 2 in Introduction). Kroll and Stewart (1994) suggest that during foreign language spoken word comprehension and production native and foreign language phonological representations are closely linked (Kroll & Stewart, 1994 - see Figure 1C in Introduction). Based on this assumption, cross-linguistic interdependence would be expected in native language spoken word comprehension and production. However, our results confirmed this prediction only for spoken word production, but not for spoken word comprehension.

Kroll and Stewart's (1994) assumption that links from native to foreign language phonological representations are weaker than connections in the opposite direction might provide a plausible explanation for our results. No clear definition of weaker and stronger links is provided within the Revised Hierarchical Model (Kroll & Stewart, 1994). Nevertheless, based on connectionist frameworks of language processing (e.g., Seidenberg, 2007), we could interpret these terms as different thresholds of activation needed to trigger activation of representations. Also taking into account Ellis and Young's (1988) model, we assume that when asked to produce a foreign spoken word to name a picture, children accessed the semantics of this word, and then the native language phonological representation from there before accessing the foreign language phonological representation. Weaker links from the native language to foreign language phonological representations would require a higher activation level to trigger successful activation from native to foreign language phonological representations. In contrast, when asked to select a picture after hearing a foreign spoken word, children accessed the phonological representation for the foreign word, then the native language phonological representation, and finally the semantics of this word. If it is true that the links from foreign to native language are stronger, lower activation levels may be sufficient to activate the foreign language first and then native language phonological representations to allow the comprehension of the spoken foreign language word. Therefore, poor native language spoken word production skills could have a larger impact on equivalent foreign language skills than poor native language spoken word comprehension skills, as observed in our results. Further research is needed to confirm this idea. Measuring response times of forward and backward translations (native to foreign language words and vice versa) with the same foreign language words used across both tasks, might allow a more direct assessment of the strength of links between native and foreign language phonological representations involved in foreign language spoken word comprehension and production (also see Kroll & Stewart, 1994).

Previous research on cross-linguistic interdependence in vocabulary skills of children with typical literacy skills has been summarized in a meta-analytic study by Melby-Lervåg and Lervåg (2011). It is, however, difficult to compare our findings to the results of this report, because receptive and expressive vocabulary measures, as well as listening comprehension tasks were combined into an overall oral language effect. Although a secondary effect was also reported for receptive vocabulary tasks only, no such information is provided for expressive vocabulary tasks. Hence, we cannot assess whether Melby-Lervåg and Lervåg's (2011) results also differed across receptive and expressive vocabulary skills, as found in this study. The information provided on receptive vocabulary skills shows a small, but significant cross-linguistic correlation which can only be interpreted with caution due to the presence of significant between study heterogeneity (e.g. significant correlation: Atwill, Burstein, Blanchard & Gorin, 2007; non-significant correlation: Carlisle, Beeman, Davis & Spharim, 1999 - see Melby-Lervåg & Lervåg's, 2011 supplemental materials for details).

For written language skills, we found cross-linguistic interdependence for word reading and spelling. According to Ellis and Young's (1988) model, successful word reading and spelling can either be achieved through sublexical or lexical processing, depending on word regularity (see Figure 2 in Introduction). While word reading and spelling in English relies more heavily on lexical processing due to a high proportion of words with irregular grapheme-phoneme correspondences, in German most words contain regular grapheme-phoneme correspondences and thus can be successfully read and spelled either through lexical or sublexical processing (Landerl, 2017).

Our results indicate a reliance on common underlying language resources for lexical processes (both word reading and word spelling). The degree to which the native language orthographic lexicon enables successful lexical processing seems to influence how well the same process will function in the foreign language. Similar results have been reported for many language combinations (e.g. Greek-English: Sotiropoulos & Hanley, 2017; Hebrew-English: Geva & Siegel, 2000; Persian-English: Gholamain & Geva, 1999). These results fit well with bilingual models of written language processing, such as the Bilingual Interactive Activation Plus model, under which the same memory system is utilised across languages and language membership is processed via tags (Dijkstra & van Heuven, 2002 - see Figure 1B in Introduction). It is likely that the presence of many cognates between German and English (e.g. <friend> in English and <Freund> in German) influences the extent to which native and foreign language lexical processing is associated (Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010; Chung, Chen & Geva, 2018; see also Bialystok, Luk & Kwan (2005) for moderating effects of language similarity on cross-linguistic interdependence in word reading skills).

However, our results indicate that native language sublexical processing (nonword reading and spelling) also plays a role in explaining individual differences in foreign language word reading and spelling. This may be expected given that even “irregular” words in English (e.g., yacht) mostly consist of regular grapheme-phoneme and phoneme-grapheme correspondences (i.e., regular graphemes: <y>, <t>), and many of these are shared between German and English (e.g. <t> = /t/ in both German and English). Sublexical contribution to lexical skills also fits with the idea that sounding out procedures facilitate the learning of lexical representations (e.g., Share, 1995; Wang, Nickels & Castles, 2014).

With regards to sublexical abilities, we first found that native language nonword spelling significantly contributed to explaining the variance observed in foreign language nonword spelling. This is in line with bilingual models of written language processing, such as the Bilingual Interactive Activation Plus model, which assumes a larger involvement of common underlying language mechanisms in sublexical processing, in contrast to language-specific mechanisms (Dijkstra & van Heuven, 2002 - see Figure 1B in Introduction). First, the general ability to map phonemes onto graphemes is a shared ability for alphabetic scripts such as in German and English (Koda, 2008). In addition, as mentioned above, German and English share many phoneme-grapheme correspondences. As a consequence, individual differences in the ability of poor readers/spellers to apply native language phoneme-grapheme conversion rules are reflected also in sublexical processing in their foreign language (e.g. German phoneme-grapheme correspondences /n/ = <n> and /m/ = <m> are likely to influence spelling of the English nonword /nu:m/) (see Yeong & Rickard Liow (2010) for similar conclusions with a population of Mandarin-English bilingual children).

In contrast, for foreign language nonword reading, we found native language word reading, instead of nonword reading to contribute to explaining the observed variability. German words are mostly regular and can therefore either be successfully read through lexical or sublexical processing. This necessarily makes our German word reading measure an index of both lexical and sublexical processing at the same time (also see influences of sublexical abilities on word reading above). Hence, it is difficult to disentangle the relative contributions of the two processes. The limitation is also statistical in nature, where multi-collinearity was observed between native language word and nonword reading (see VIF values reported in results section). As a consequence, we likely do not have enough power to detect a significant contribution of native language nonword reading to explain individual differences in the same foreign language task.

Other cross-linguistic studies of bilingual children with many different language combinations have found that native and foreign language sublexical processing rely on common underlying mechanisms (e.g. Arabic-English: Abu-Rabia & Siegel, 2002; Portuguese-English: Da Fontoura & Siegel, 1995; Italian-English: D'Angiulli, Siegel & Serra, 2001). It seems likely that our data are best interpreted as showing the same for German children learning English, but further studies will be needed to confirm this interpretation.

In addition to finding cross-linguistic interdependence between native and equivalent foreign language subskills in children with poor literacy skills, our results also revealed cross-linguistic associations between native language word reading and foreign language spoken word production skills, as well as the opposite pattern. These results could be a consequence of the specific foreign language vocabulary instruction received by the participants of this study. As foreign language teachers reported, children were asked to memorize new foreign language words from written vocabulary lists included in their textbooks. In this material, each new foreign language word was presented together with its native language translation. Individual differences in native language word reading skills are therefore required to access spoken foreign language word forms.

Lastly, as a secondary aim we explored the role of broader cognitive measures as potential sources of individual variability in foreign language attainment of poor readers/spellers. We found foreign language learning motivation, but not linguistic background, intellectual ability, short term memory or working memory to contribute to explaining the variability observed in foreign language attainment. Further, foreign language learning motivation only played a significant role in explaining individual differences in foreign language spoken word production. Children with higher foreign language learning motivation might dedicate more time to memorizing new vocabulary items, which in turn improves their spoken word production skills. This might be especially true in the context of the specific foreign language vocabulary instruction methodology mentioned above.

However, it remains unclear if foreign language learning motivation arises as a consequence of successful foreign language attainment or if the opposite is true (Li, 2016; Sparks & Ganschow, 1991). While studies such as ours can identify associations between different skills, longitudinal or training studies are better suited to explore the direction of any potential causal links in this respect. Overall, our results emphasize the important role of native language skills in contrast to broader cognitive measures to explain individual differences in foreign language attainment of poor readers/spellers.

Conclusions

To conclude, the current study integrated contributions from current models of bilingual processing into a well-known language processing model with a longstanding history of being applied in clinical settings to diagnose language deficits across different languages (Friedmann et al., 2013; Friedmann & Coltheart, 2016; Gray & Kiran, 2013; Kezilas et al., 2014; Kohnen et al., 2012; Kohnen et al., 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003). From a theoretical perspective, this study therefore provides a framework for analysing cross-linguistic interdependence at different levels of language processing. We proposed this framework to disentangle the influence of language-common and language-specific resources at input and output, as well as during oral and written levels of cross-linguistic processing in children with poor literacy skills. However, it might ultimately also be of practical use for teachers and clinicians working with other bilingual populations with language disorders.

Furthermore, to the best of our knowledge, this study is unique in investigating if individual differences in foreign language attainment of children with poor literacy skills are associated with individual differences in corresponding native language skills. Our findings indicate that poor readers/spellers with poor native language speech sound perception, spoken word production, word reading and spelling and nonword spelling skills might be at greater risk of experiencing the same foreign language deficits. As our results only point to associations between native and foreign language skills, future longitudinal and training studies are needed to confirm the presence and direction of causal links between individual differences in native and foreign language skills.

References

Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of psycholinguistic research*, *31*(6), 661-678. doi: 10.1023/A:1021221206119

Atwill, K., Blanchard, J., Gorin, J. S., & Burstein, K. (2007). Receptive vocabulary and cross-language transfer of phonemic awareness in kindergarten children. *The Journal of Educational Research*, *100*(6), 336-346. doi: 10.3200/JOER.100.6.336-346

Baayen, R. H. (2008). *Analysing linguistic data: A practical introduction to statistics using R.* Cambridge: University Press.

Baayen, R.H., Piepenbrock, R., & van RIJN, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baker, C. (1996). *Foundations of Bilingual Education and Bilingualism*. (2.ed.) UK: Multilingual Matters Ltd.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4.* R package version, 1(7), 1-23. Retrieved from http://CRAN.R-project.org/package=lme4.b.

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (Eds.) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific studies of reading*, *9* (1), 43-61. doi: 10.1207/s1532799xssr0901_4.

Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*(2), 131-158.

Bohn, O. S., Best, C. T., Avesani, C., & Vayra, M. (2011). Perceiving through the lens of native phonetics: Italian and Danish listeners' perception of English consonant contrasts. *International Congress of Phonetic Sciences*, 17, 336-339.

Carlisle, J. F., Beeman, M., Davis, L. H., & Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics*, *20*(4), 459-478. doi: 10.1017/S0142716499004014

Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*, 46(1), 71-86. doi: 10.1017/S0261444811000218.

Chung, S. C., Chen, X., & Geva, E. (2018). Deconstructing and reconstructing cross-language transfer in bilingual reading development: An interactive framework. *Journal of Neurolinguistics*. doi: 10.1016/j.jneuroling.2018.01.003.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108* (1), 204-256. doi: 10.1037/0033-295X.108.1.204

Csizér, K., & Magid, M. (Eds.). (2014). *The impact of self-concept on language learning*. UK: Multilingual Matters Ltd.

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of cross-cultural psychology*, 9(2), 131-149.

Cummins, J. (1989). A theoretical framework for bilingual special education. *Exceptional children*, 56(2), 111-119.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. UK: Multilingual Matters Ltd.

D'Angiulli, A., Siegel, L., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics*, *22*(4), 479-507. doi: 10.1017/S0142716401004015

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. doi: 10.1007/BF01026951

Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197. doi: 10.1017/s1366728902003012

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

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.

Ellis, A.W. & Young, A.W. (1988) *Human cognitive neuropsychology*. Hove: Erlbaum.

Flege, J.E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp.233-277). Timonium, MD: York Press.

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. In A. Baron, & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives*. Berlin, Boston: De Gruyter Mouton.

Friedmann, N., Biran, M. & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K.K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp.350-374). Cambridge University Press.

Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German-and English-speaking children. *Scientific Studies of reading*, *2*(1), 31-54.

Gardner, R. C. (1985). *The attitude/motivation test battery: Technical report*. University of Western Ontario. Retrieved from http://publish.uwo.ca/~gardner/docs/AMTBmanual.pdf.

Genesee, F. (2015). Myths About Early Childhood Bilingualism. *Canadian Psychology,* 56(1), 6-15. doi: 10.1037/a0038599.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing*, *12*(1-2), 1-30. doi: 10.1023/A:1008017710115

Geva, E., Wade-Woolley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in two different orthographies. *Journal of Reading Behaviour*, *25*(4), 383-406.

Gholamain, M., & Geva, E. (1999). Orthographic and Cognitive Factors in the Concurrent Development of Basic Reading Skills in English and Persian. *Language Learning*, *49*(2), 183-217. doi: 10.1111/0023-8333.00087

Glück, C. (2011). *Wortschatz- und Wortfindungstest für 6- bis 10-jährige (WWT 6-10).* [Vocabulary and word-retrieval test for 6-10 year old children] Munich: Elsevier, Urban & Fischer.

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). *Language acquisition across linguistic and cognitive systems*, *52*, 267-283. doi: 10.1075/lald.52.18gra.

Gray, T., & Kiran, S. (2013). A theoretical account of lexical and semantic naming deficits in bilingual aphasia. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1314-1327. doi: 10.1044/1092-4388(2012/12-0091).

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. Language and Cognitive Processes, 3, 233–274. doi: 10.1080/01690968808402089.

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A.M.B. de Groot, J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, (pp. 225-254). NJ: Erlbaum.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and cognition*, 1(2), 131-149. doi:10.1017/S136672899800025X.

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

Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of Psycholinguistic Research*, 34(6), 603-618. doi: 10.1177/0022219409345018.

Holm, A. and Dodd, B. (2006). Phonological development and disorder of bilingual children acquiring Cantonese and English. In Z. Hua and B. Dodd (Ed.), *Phonological Development and Disorders in Children. A multilingual perspective* (pp. 286-325) Clevedon: Multilingual Matters.

Hussien, A. M. (2014). The effect of learning English (FL) on learning of Arabic literacy (NL) in the primary school. *International Education Studies*, 7(3), 88-98. Retrieved from http://www.ccsenet.org/journal/index.php/ies/article/viewFile/33294/19765.

Katz, L. & Frost, R. (1992). *The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis*. In R. Frost & L. Katz (Eds.). *Orthography, Phonology, Morphology, and Meaning* (pp. 67–84). Amsterdam: Elsevier North Holland Press.

Kezilas, Y., Kohnen, S., McKague, M., & Castles, A. (2014). The locus of impairment in English developmental letter position dyslexia. *Frontiers in human neuroscience*, *8*, 356. doi: 10.3389/fnhum.2014.00356

Koda, K. (2008). I*mpacts of Prior Literacy Experience on Second-Language Learning to Read*. In K. Koda, & A. M. Zehler (Eds.), Learning to Read across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development (pp. 68-96). Mahwah, NJ: Erlbaum.

Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in Grades 1-7. *Australian Journal of Learning Difficulties*, 20(1), 15-38. doi:10.1080/19404158.2015.1023209

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime' becomes 'smile': Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681–3692. doi:10.1016/j.neuropsychologia.2012.07.016

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex*. doi: 10.1016/j.cortex.2018.04.008

Kohnert, K., Windsor, J., & Miller, R. (2004). Crossing borders: Recognition of Spanish words by English-speaking children with and without language impairment. *Applied Psycholinguistics*, *25*(4), 543-564. doi: 10.1017/S0142716404001262

König, E. & Gast, V. (2012). *Understanding English-German Contrasts*. Berlin: Erich Schmidt Verlag.

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

Landerl, K. (2017). Learning to Read German. In L. Verhoeven & C. Perfetti (2017). *Learning to Read across Languages and Writing Systems* (pp. 299-322). United Kingdom: Cambridge University Press.

Landerl, K., Wimmer, H., & Frith, U. (1996). The impact of orthographic consistency on dyslexia: A German-English comparison. *Cognition*, *63*(3), 315-334. doi: 10.1016/S0010-0277(97)00005-X.

Lenhard, W. & Schneider, W. (2006*). ELFE 1-6: ein Leseverständnistest für Erst-bis Sechstklässler* [A reading comprehension test for 1st to 6th graders]. Göttingen: Hogrefe.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38(4), 801-842. doi: 10.1017/S027226311500042X.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic bulletin & review*, 21(4), 861-883. doi: 10.3758/s13423-013-0565-2.

May, P. (2012). *Die Hamburger Schreibprobe (HSP)* [The Hamburger Writing Probe]. Stuttgart: VPM.

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. doi: 10.1111/j.1467-9817.2010.01477.x

Moll, K. & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest; Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT)* [SLRT-II: Reading and spelling test; Revision of the Salzburger Reading and spelling test (SLRT)]. Bern: Huber.

Montessori, M. (2013). *The Montessori Method*. New Brunswick: Transaction publishers.

Müller, F. (2010). Verteilung von Armut im Primarbereich in Berlin. [Distribution of poverty in Primary schools in Berlin]. *Zeitschrift für Inklusion* [Journal for Inclusion]. Retrieved from http://www.inklusion-online.net/index.php/inklusion-online/article/view/110/110.

Palladino, P., Bellagamba, I., Ferrari, M. & Cornoldi, C. (2013). Italian Children with Dyslexia are also poor in reading English words, but accurate in reading English pseudowords. *Dyslexia* 19, 165-177. doi: 10.1002/dys.1456.

Paradis, J., & Jia, R. (2017). Bilingual children's long-term outcomes in English as a second language: language environment factors shape individual differences in catching up with monolinguals. *Developmental science*, 20(1). doi: 10.1111/desc.12433.

Peer, L. & Reid, G. (2014). *Multilingualism, literacy & dyslexia*. UK: Taylor & Francis Group.

Petermann & Petermann (2014). *Wechsler Intelligence Scale fro Children - Fourth Edition (WISC IV) - Deutschsprachige Adaptation nach D. Wechsler* [German adaptation after D. Wechsler]. Frankfurt a. M.: Pearson.

Proctor, C. P., August, D., Snow, C., & Barr, C. D. (2010). The interdependence continuum: A perspective on the nature of Spanish–English bilingual reading comprehension. *Bilingual Research Journal*, *33*(1), 5-20. doi: 10.1080/15235881003733209.

Seidenberg, M.S. (2007). Connectionist models of reading. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (p. 235-250). Oxford: Oxford University Press.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151-218. doi: 10.1016/0010-0277(94)00645-2

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two tongues, one brain: imaging bilingual speech production. *Frontiers in Psychology*, *2*, 166. doi: 10.3389/fpsyg.2011.00166.

Soroli, E., Szenkovits, G., & Ramus, F. (2010). Exploring dyslexics' phonological deficit III: foreign speech perception and production. *Dyslexia*, *16*(4), 318-340. doi: 10.1002/dys.415

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216. doi: 10.1016/j.cognition.2017.06.024

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreign Language Annals*, 49(2), 252-270. doi:10.1111/flan.12196.

Sparks, R. L., & Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *The Modern Language Journal*, 75(1), 3-16. doi:10.2307/329830.

Sparks, R. L., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992a). Identifying native language deficits in high- and low-risk foreign language learners in high school. *Foreign Language Annals*, 25(5), 403.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia*, 56(1), 129-160. doi: 10.1007/s11881-006-0006-2.

Sparks, R., Ganschow, L., & Pohlman, J. (1989). Linguistic coding deficits in foreign language learners. *Annals of Dyslexia*, 39(1), 177-195. doi:10.1007/BF02656908.

Sparks, R., Ganschow, L., Javorsky, J., Pohlman, J., & Patton, J. (1992b). Test Comparisons among Students Identified as High-Risk, Low-Risk, and Learning Disabled in High School Foreign Language Courses. *The Modern Language Journal*, 76(2), 142-159. doi:10.2307/329768.

Stadie, N. & Schröder, A. (2009). Kognitiv orientierte Sprachtherapie [Cognitive oriented speech therapy]. Munich: Elsevier.

Stadie, N., & van de Vijver, R. (2003). A linguistic and neuropsychological approach to remediation in a German case of developmental dysgraphia. *Annals of Dyslexia*, *53*(1), 280-299.

Stadie, N., Cholewa, J. & De Bleser, R., (2013). *LeMo - Lexikon modellorientiert. Diagnostik bei Aphasie, Dyslexie und Dysgraphie* [LeMo - Model oriented lexicon. Diagnostic of aphasia, dyslexia and dysgraphia]. Hofheim: NAT-Verlag.

Steiner, R. (1997). *The essentials of education*. Hudson: SteinerBooks.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development,* 47, 466-472.

van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of Learning Disabilities*, 39(1), 74-90. doi:10.1177/00222194060390010701.

Verhoeven, L. T. (1994). Transfer in bilingual development: The linguistic interdependence hypothesis revisited. *Language learning*, 44(3), 381-415. doi:10.1111/j.1467-1770.1994.tb01112.x.

von Hagen, A., Kohnen, S. & Stadie, N. (submitted). Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis.

von Hagen, A., Kohnen, S. & Stadie, N. (to be submitted). Individual differences in foreign language attainment of children with poor literacy skills.

Wang, H., Nickels, L. & Castles, A. (2015) Orthographic learning in developmental surface and phonological dyslexia, *Cognitive Neuropsychology*, 32:2, 58-79, doi: 10.1080/02643294.2014.1003536.

Wight, M.C.S. (2015). Students With Learning Disabilities in the Foreign Language Learning Environment and the Practice of Exemption. *Foreign language Annals*, 48(1), 39-55. doi: 10.1111/flan.12122.

Yap, M. & Rickard Liow, S. (2016). *Processing the written word*. In V. Cook and D. Ryan (Eds.) The Routledge Handbook of the English Writing System (pp. 457-474). New York: Routledge.

Yeong, S. H., & Rickard Liow, S. (2010). Phonemic representation and early spelling errors in bilingual children. *Scientific Studies of Reading*, 14(5), 387-406. doi: 10.1080/10888431003623538.

Supplemental materials

Nonword discrimination and repetition tasks.

Table 1

Contrasts between phonological systems of German and English

	German	English
Language specific phoneme	ç , x	θ , ð
	ʁ	ɹ
	t͡s	w
	øː , œ	ɜː
	yː , Y	æ
Phonetic difference	final devoicing of voiced obstruents -d, -g, -b	velarized syllable-final -ɫ
	initial cluster ʃl- , ʃm- , ʃp-	initial cluster sl- , sm- , sp-
Phonotactic difference	initial cluster plosive + nasal kn-, gn-, pn-	Initial cluster consonant + glide mj-, lj-, fj-

Note. This table was adapted from König & Gast (2012)

Table 2

Native language nonword discrimination task.

Item	Minimal pair	Item A	Item B	Native language specific target phoneme	Native and foreign language common contrast phoneme
1	no	kɪç	kɪʃ	ç	ʃ
2	yes	kɪç	kɪç	ç	ʃ
3	no	pʊx	pʊk	x	k
4	yes	pʊx	pʊx	x	k
5	no	tsʊp	zʊp	ts	z
6	yes	tsʊp	tsʊp	ts	z
7	no	bø:n	bo:n	ø:	o:
8	yes	bø:n	bø:n	ø:	o:
9	no	pœs	pɔs	œ	ɔ
10	yes	pœs	pœs	œ	ɔ
11	no	my:ʃ	mʉ:ʃ	y:	ʉ:
12	yes	my:ʃ	my:ʃ	y:	ʉ:
13	no	gYf	gʊf	Y	ʊ
14	yes	gYf	gYf	Y	ʊ
15	no	kni:k	ni:k	kn-	n-
16	yes	kni:k	kni:k	kn-	n-
17	no	gni:m	ni:m	gn-	n-
18	yes	gni:m	gni:m	gn-	n-
19	no	pnɪf	nɪf	pn-	n-
20	yes	pnɪf	pnɪf	pn-	n-

Note. Items were presented in a randomised order.

Table 3

Native language nonword repetition task.

Item	Target phoneme	Syllable length	IPA transcription
1	ç	1	pi:ç
2	ç	2	ʃi:pɪç
3	x	3	fɪbətʊx
4	ʁ	1	ʁɪm
5	ʁ	2	gi:ʁɪp
6	ʁ	3	mɪbəɹɪk
7	ts	1	tsʊp
8	ts	2	di:tsɪm
9	ts	3	gʊbətsɪm
10	øː	1	møːʃ
11	ø:	2	hi:løːt
12	ø:	3	hiːfətøːt
13	Y	1	dYt
14	Y	2	fi:pYm
15	Y	3	liːdəfYp
16	-l	1	ʃɪl
17	-l	2	ʃi:təl
18	-l	3	pi:fədʊl
19	ʃl	1	ʃli:n
20	ʃp	2	bi:ʃpɪk
21	ʃm	3	vi:dəʃmɪp
22	kn-	1	kni:k
23	gn-	2	bi:gnɪf
24	pn-	3	hi:dəpnɪf

Note. Items were presented in a randomised order.

Parent and student questionnaires.

Figure 1 Parent questionnaire.

Research project 'Dyslexia and second language learning' – Potsdam University

Parent-questionnaire

Code number:

Please fill in this questionnaire carefully and return it together with the signed consent form to your child´s teacher. Participation in this questionnaire is voluntary and the provided information will be treated confidentially. Thank you very much!

Child´s date of birth: ____________ Gender: ☐ male ☐ female

In which pregnancy week was your child born? _______________________

Does your child have any loss of hearing? ☐ No ☐ Yes

If yes, please give details: _____________________________________

Does your child have any vision impairments? ☐ No ☐ Yes

If yes, please give details :______________________________________

Which language is your child exposed to at home ?

☐ Only/ mainly German.

☐ German and another language in an equal proportion. Which other language(s)? __________

☐ Mainly another language than German. Which other language(s)? ________________

Does your child have any developmental difficulties? No ☐ Yes ☐

If yes, which?

☐ Reading or spelling disorder (dyslexia or dysgraphia)

☐ Specific Language Impairment (SLI)

☐ Math learning difficulty (Dyscalculia)

☐ Attentional deficit disorder (ADD)

☐ Difficulties in second language learning

☐ Developmental coordination disorder (dyspraxia)

☐ Other: ________

Does a close relative of your child (siblings, parents, grandparents) have developmental difficulties?

☐ No ☐ Yes

If yes, which?

☐ Reading or spelling disorder (dyslexia or dysgraphia)

☐ Specific Language Impairment (SLI)

☐ Math learning difficulty (Dyscalculia)

☐ Attentional deficit disorder (ADD)

☐ Difficulties in second language learning

☐ Developmental coordination disorder (dyspraxia)

☐ Other: _________

Did or does your child participate in an intervention with a speech therapist, psychologist or specialised teacher? No ☐ Yes ☐

If yes, what was the main focus of this intervention?

☐ Speech perception and production		☐ Vocabulary
☐ Grammar	☐ Reading	☐ Writing
☐ Math	☐ Second language learning	☐ Attention
☐ Behaviour	☐ Other:____________	

This questionnaire was completed by: ☐ mother ☐ father

☐ other person: ________________________

Thank you very much!

Kontakt:

MSc Alexa von Hagen
Department Linguistik
Universität Potsdam
Karl-Liebknecht-Str. 24-25
14476 Potsdam
Tel.: +49-015226148810
E-mail: alexavh@gmail.com

Dr. phil. Nicole Stadie
Department Linguistik
Universität Potsdam
Karl-Liebknecht-Str. 24-25
14476 Potsdam
Tel.: +49331/9772935
E-mail: nstadie@uni-potsdam.de
www.uni-potsdam.de/treatmentlab

Note. The original questionnaire was presented in German.

Figure 2 Student questionnaire.

Code number: Date:

Student questionnaire

1. Which languages do you speak where and with whom?		
⊗ German	○ at home ○ with friends	⊗ at school ○ with:__________________
⊗ English	○ at home ○ with friends	⊗ at school ○ with:__________________
○ another language:____________	○ at home ○ with friends	○ at school ○ with:__________________
○ another language:____________	○ at home ○ with friends	○ at school ○ with:__________________

2. Which of the languages you speak do you understand best and are able to speak in best?

1. Place: ____
2. Place: ____
3. Place: ____

3. In which language can you read and write the best?

1. Place: ____
2. Place: ____
3. Place: ____

4. Have you ever lived in another country? ○ **Yes** ○ **No**

If yes, in which country/countries? ______________________

If yes, how old were you, when you arrived there? ○ I was born there. ○___ years old.

If yes, how old were you, when you left that country/countries? ___ years old.

5. When did you start receiving German instruction?

○ before grade 1 ○ in grade 1 ○ in grade 2 ○ in grade 3

○ in grade 4 ○ in grade 5 ○ in grade 6

Note. The original questionnaire was presented in German.

Foreign language learning motivation questionnaire.

Figure 1 Questionnaire adapted from Gardner (1985).

SG1

Code number:

Motivation questionnaire

	X			?	✓		
	I don't agree			I don't know	I agree		
	not at all	more or less	a little bit		a little bit	more or less	totally
	1	2	3	4	5	6	7
Playing soccer is fun.							

	X			?	✓		
	I don't agree			I don't know	I agree		
	not at all	more or less	a little bit		a little bit	more or less	totally
1.English is an important school subject.	1	2	3	4	5	6	7
2. I hate learning English.	1	2	3	4	5	6	7
3. I am planning on learning as much English as possible.	1	2	3	4	5	6	7
4. When I finish school, I won´t learn English anymore because it doesn't interest me.	1	2	3	4	5	6	7
5. I love learning English.	1	2	3	4	5	6	7
6. I would prefer spending more time on other subjects than English.	1	2	3	4	5	6	7
7. Learning English is important because it helps you communicate with people who speak that language.	1	2	3	4	5	6	7
8. Learning English is important, because it allows you to understand music, films, games and books in that language.	1	2	3	4	5	6	7
9. Learning English is important because you widen your general knowledge this way.	1	2	3	4	5	6	7
10. Learning English is important because it might give you better job opportunities.	1	2	3	4	5	6	7

	X			?	✓		
	I don't agree			I don't know	I agree		
	not at all	more or less	a little bit		a little bit	more or less	totally
11. I am embarrassed to speak English in class.	1	2	3	4	5	6	7
12. I believe that my classmates speak better English than I do.	1	2	3	4	5	6	7
13. I get nervous and make mistakes when I have to participate in English class.	1	2	3	4	5	6	7
14. I am afraid that my classmates might laugh when I have to say something in English class.	1	2	3	4	5	6	7

15. If I didn't have English classes at school, I would:
   a. learn the language in other ways (through English songs, films, etc.).
   b. not learn English at all.
   c. try to receive English classes somewhere else.

16. When I don't understand something in English class:
   a. I ask our teacher immediately.
   b. I ask just before I have to take a test.
   c. I don´t worry about it.

17. If I'm totally honest, I my effort in English class is:
   a. just enough to follow the class.
   b. very low because I get a good grade anyway.
   c. very high because I am trying to learn to speak English as well as I can.

18. When it comes to speaking English in class:
   a. I participate happily as much as I can.
   b. I only answer easy questions.
   c. I never say anything.

19. If I had the possibility to watch TV in English, I would:
   a. never do that.
   b. do it sometimes.
   c. do it as much as possible.

20. When I hear somebody talking in English,
   a. I try to understand some of the easy words.
   b. I try to understand everything.
   c. I don't even listen.

Note. The original questionnaire was presented in German.

Foreign language nonword discrimination and repetition

Table 1

Foreign language nonword discrimination task.

Item	Minimal pair	Item A	Item B	Foreign language specific target phoneme	Native and foreign language common contrast phoneme
1	yes	θi:p	ti:p	θ	t
2	no	θi:p	θi:p	θ	θ
3	yes	θɪm	fɪm	θ	f
4	no	θɪm	θɪm	θ	θ
5	yes	dʊθ	dʊs	θ	s
6	no	dʊθ	dʊθ	θ	θ
7	yes	ði:t	di:t	ð	d
8	no	ði:t	ði:t	ð	ð
9	yes	ðɪm	vɪm	ð	v
10	no	ðɪm	ðɪm	ð	ð
11	yes	ðʊk	zʊk	ð	z
12	no	ðʊk	ðʊk	ð	ð
13	yes	wɪf	vɪf	w	v
14	no	wɪf	wɪf	w	w
15	yes	wi:m	li:m	w	l
16	no	wi:m	wi:m	w	w
17	yes	wi:f	ji:f	w	j
18	no	wi:f	wi:f	w	w
19	yes	pæf	pɛf	æ	ɛ
20	no	pæf	pæf	æ	æ
				Foreign language specific target phoneme	Native language specific contrast phoneme
21	no	ɹʊp	ʁʊp	ɹ	ʁ
22	yes	ɹʊp	ɹʊp	ɹ	ɹ
23	no	spɪf	ʃpɪf	sp-	ʃp-
24	yes	spɪf	spɪf	sp-	sp-
25	no	nɜːk	nøːk	ɜː	øː
26	yes	nɜːk	nɜːk	ɜː	ɜː

Note: Items were presented in a randomised order.

Table 2

Foreign language nonword repetition task.

Item	Target phoneme	Syllable length	IPA transcription
1		1	θiːp
2	θ and ð	2	biːðɪm
3		3	mɪbəθiːp
4		1	ɹɪʃ
5	ɹ	2	giːɹɪp
6		3	mɪbəɹɪk
7		1	wiːf
8	w	2	diːwif
9		3	biːzəwɪm
10		1	nɜːk
11	ɜː	2	hiːlɜːt
12		3	hiːfəgɜːt
13		1	dæt
14	æ	2	hɪpæf
15		3	tiːdəfæp
16		1	pɪd
17	-d, -g, -b	2	miːlɪg
18		3	piːfədʊb
19		1	sliːʃ
20	sp-, sl-, sm-	2	beespɪk
21		3	ʃiːdɪsmət
22		1	mjʊə
23	mj, lj, fj	2	hɪljʊə
24		3	ʃiːvəfjʊə

Note: Items were presented in a randomised order.

Foreign language spoken word-picture matching and picture naming tasks.

Table 1

Foreign language vocabulary test items.

Item	Word class	Time of acquisition	N of syllables	Presence of cluster	Cognate status	Item
1				cluster	cognate	hand
2			1 syllable		noncognate	plate
3				no cluster	cognate	cow
4		early acquired (3rd & 4th grade)			noncognate	girl
5				cluster	cognate	apple
6			2-3 syllables		noncognate	flower
7				no cluster	cognate	banana
8	Noun				noncognate	window
9				cluster	cognate	ghost
10			1 syllable		noncognate	spoon
11				no cluster	cognate	fire
12		late acquired (5th & 6th grade)			noncognate	bag
13				cluster	cognate	elephant
14			2-3 syllables		noncognate	vegetables
15				no cluster	cognate	calendar
16					noncognate	cinema
17				cluster	cognate	blue
18			1 syllable		noncognate	small
19				no cluster	cognate	red
20		early acquired (3rd & 4th grade)			noncognate	sad
21				cluster	cognate	orange
22			2-3 syllables		noncognate	cloudy
23				no cluster	cognate	sunny
24	Adjective				noncognate	yellow
25				cluster	cognate	brown
26			1 syllable		noncognate	fast
27				no cluster	cognate	long
28		late acquired (5th & 6th grade)			noncognate	wrong
29				cluster	cognate	hungry
30			2-3 syllables		noncognate	angry
31				no cluster	cognate	open
32					noncognate	empty
33						fly
34					cognate	help
35						sleep
36	Verb	early acquired (3rd & 4th grade)	1-2 syllables	cluster		swim
37						draw
38					noncognate	play
39						climb

40	late acquired (5th & 6th grade)	no cluster	cognate	travel
41				sing
42				learn
43				run
44				wash
45			noncognate	read
46				write
47				catch
48				choose

Table 2

Coding scheme used to determine cognate status of foreign language vocabulary items following Kohnert, Windsor & Miller (2004, p.524).

Feature Overlap	Scoring	Example
Initial sound	3 = Same consonant	banana – Banane
(0-3 points)	2 = Same vowel	open - offen
	1= Similar sound (e.g., same sound class or one of consonant cluster)	swim - schwimmen
	0 = Complete mismatch between initial sounds	small – klein
Number of syllables	2 = Same number of syllables	cow - Kuh
(0-2 points)	1 = Different by only 1 syllable	fly - fliegen
	0 = Different by more than 1 syllable	choose – aussuchen
Consonants	3 = >70% consonant overlap	fire - Feuer
(0-3 points)	2 = 50-70% consonant overlap	red - rot
	1 = <50% consonant overlap	play - spielen
	0 = No consonant overlap	sad – traurig
Vowels	2 = ≥80% vowel overlap	elephant - Elefant
(0-2 points)	1 = 50-80% vowel overlap	blue - blau
	0 = no vowel overlap	cloudy - windig

Note. This coding scheme was developed following Kohnert et al. (2004, p.524).

Each English stimulus word was scored relative to its German translation equivalent on the four different features, for a maximum score of 10. The items that presented a score from 0-4 points were considered non-cognates, while 5-10 points corresponded to a cognate status.

Foreign language reading and spelling tasks.

Table 1

Foreign language irregular word reading and spelling

Item nr.	Time of exposure	Cognate status	Presence of cluster	Item
1	early exposure (3rd & 4th grade)	cognate	cluster	friend
2				cold
3				month
4			no cluster	live
5				shoe
6				house
7		noncognate	cluster	climb
8				small
9				great
10			no cluster	put
11				head
12				eye
13	late exposure (5th & 6th grade)	cognate	cluster	break
14				group
15				find
16			no cluster	love
17				fall
18				come
19		noncognate	cluster	fast
20				ask
21				should
22			no cluster	key
23				choose
24				get

Table 2

Foreign language nonword reading and spelling

		Reading				Spelling			
Item nr.	GPC[a]/PGC[b] contrast	Item	Expected pronunciation	English GPC	Conflicting German GPC	Item	Expected spelling[c]	English PGC	Conflicting German PGC
1	new FL phoneme – common NL/FL grapheme	lan	læn	<a> = /æ/	<a> = /a/	læn	lan	/æ/ =<a>	/æ/ = ∅
2		baf	bæf			bæf	baf		
3		mur	mɜ:r	<ur> = /ɜː/	<ur> = /uə/	mɜ:r	mur (mir, mer)	/ɜː/ = <ur>	/ɜː/ = ∅
4		hurm	hɜ:m			hɜ:m	hurm (hirm)		
5		thix	θ iks	<th> = /θ/	<th> = /t/	θ iks	thix (thicks, thiks)	/θ/ = <th>	/θ/ = ∅
6		theck	θ ek			θ ek	theck (thek)		
7		wip	wɪp	<w> = /w/	<w> = /v/	wɪp	wip (whip)	/w/ = <w>	/w/ = ∅
8		wiss	wɪs			wɪs	wiss (wis, whiss, whis)		
9	new FL grapheme – common NL/FL phoneme	shim	ʃɪm	<sh> = /ʃ/	<sh> = ∅	ʃɪm	shim	/ʃ/ = <sh>	/ʃ/ = <sch>
10		shep	ʃep			ʃep	shep		
11		noom	nu:m	<oo> = /uː/	<oo> = /oː/	nu:m	noom	/uː/ = <oo>	/uː/ = <u>
12		doot	du:t			du:t	doot		
13		knep	nep	<n> = /kn/	<kn> = /kn/	nep	knep (nep)	/n/ = <n>	/n/ = <n>
14		knoff	nɔf			nɔf	knoff (nof, knof, noff)		
15	common NL/FL phoneme and grapheme, but new correspondence	bleen	bli:n	<ee> = /iː/	<ee> = /eː/	bli:n	bleen (blean, blene)	/iː/ = <ee>	/iː/ = <ie>
16		deeks	di:ks			di:ks	deeks (decks, deaks, deacks, deex, deax)		
17		jeck	ʤek	<j> = /ʤ/	<j> = /j/	ʤek	jeck (jek, geck, gek)	/ʤ/ = <j>	/ʤ/ = <dsch>
18		jit	ʤɪt			ʤɪt	jit (git)		
19		zill	zɪl	<z> = /z/	<z> = /ʦ/	zɪl	zill	/z/ = <z>	/z/ = <s>
20		zem	zem			zem	zem		
21		chim	ʧɪm	<ch> = /ʧ/	<ch> = /ç/	ʧɪm	chim	/ʧ/ = <ch>	/ʧ/ = <tsch>
22		chet	ʧet			ʧet	chet		
23		yoll	jɔl	<y> = /j/	<y> = /i/	jɔl	yoll (yol)	/j/ = <y>	/j/ = <j>
24		yem	jem			jem	yem		

Note. [a]GPC= Grapheme-phoneme correspondence; [b]Phoneme-grapheme correspondence; [c]We listed alternative spellings in brackets. These alternatives were accepted as correct based on the following criteria previously used by Kohnen et al. (2015): only PGC with a type frequency of 20 and a token frequency of at least 20000 out of all words in the CELEX database (Baayen et al., 1993) were scored

Chapter 4

Lexical quality in foreign language word knowledge of children with poor literacy skills

This chapter has been prepared to be submitted as:

von Hagen, A., Stadie, N. & Kohnen, S. (to be submitted). Lexical quality in foreign language word knowledge of children with poor literacy skills.

Abstract

This study aimed to investigate lexical quality in foreign language word knowledge of children with poor native language reading and spelling skills. More specifically, we asked (1) how variable the performance of individual children is across different tasks, (2) in which foreign language tasks children show the strongest and the weakest performance and (3) if different word characteristics influence foreign language word knowledge. Ten English speaking children with poor literacy skills who were learning German as a foreign language were asked to complete six different tasks involving the same 48 German words. Results revealed that (1) the quality of foreign language lexical representations varied across individuals and words, (2) the weakest performance was binding from phonological to orthographic word knowledge (measured by word spelling) for most poor readers/spellers and (3) grammatical class, cognate status, presence of consonant clusters and syllable length impact foreign language word knowledge in children with poor literacy skills.

Keywords: dyslexia, literacy skills, foreign language learning, bilingualism, second language learning

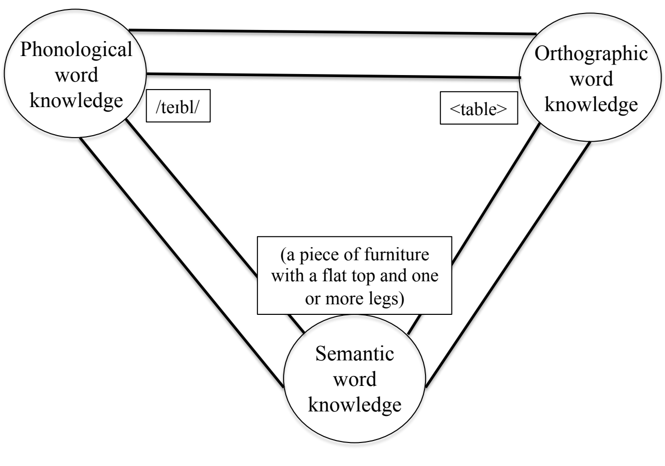
Introduction

Children identified with poor literacy skills in their native language are often thought to experience more difficulties in learning a foreign language than their classmates with typical literacy skills (Sparks, 2016; Wight, 2015). However, past research has shown that the success achieved by poor readers/spellers when learning a foreign language is variable (e.g., Helland & Kaasa, 2005; van der Leij & Morfidi, 2006; von Hagen, Kohnen & Stadie, in preparation; von Hagen, Kohnen & Stadie, under review). Some poor readers/spellers fall behind while others are just as successful as their classmates. Individual differences have also been identified with regards to what exact foreign language skills are difficult for poor readers/spellers (Morfidi, van der Leij, de Jong, Scheltinga & Bekebrede, 2007; Palladino, Bellagamba, Ferrari & Cornoldi, 2013; von Hagen et al., in preparation; von Hagen et al., under review). Indeed, when foreign language difficulties are present, poor readers/spellers seem to struggle only in some subskills (e.g. producing spoken words, reading and spelling words), but achieve average-like performances in others (e.g. comprehending spoken words) (Helland & Kaasa, 2005; Morfidi et al., 2007; Palladino et al., 2013; van der Leij & Morfidi, 2006; von Hagen et al., in preparation).

Among the different competencies that have to be mastered in foreign language learning, word knowledge plays a central role. Comprehending and producing oral and written words are foundational skills to develop more complex aspects of language, such as sentence and text comprehension and production (Nation, 2013). Thus, many activities in the foreign language classroom target word knowledge. A prominent framework to investigate individual differences in native language word knowledge is the lexical quality hypothesis proposed by Perfetti and Hart (2002) (see also Perfetti, 2007 and 2017 for updated versions). Although this framework refers to native language word knowledge, we believe it to also be suitable to investigate foreign language word knowledge.

Lexical quality refers to the extent to which a word can be successfully used across different tasks. Words with high quality representations enable a reliable and coherent performance in accessing different aspects of the word, such as its pronunciation, meaning and written form. In contrast, words with low quality representations are associated with inconsistent performance in which, for example, the meaning of a word can be accessed, but its pronunciation or written form cannot. According to this framework, the quality of lexical representations is a result of the specification of so-called phonological, orthographic and semantic constituents of the word. For example, a high quality representation of the word *table*, would enable accurate retrieval of the phonological (/teɪbl/), orthographic (<table>) and semantic (e.g. a piece of furniture with a flat top and legs) information. In addition, in high quality representations these three constituents are tightly linked together reflecting 'strong constituent binding'. In contrast, low quality lexical representations have underspecified constituents and weak constituent binding. Moreover, the quality of lexical representations is assumed to vary across readers and across different words. For instance, poor readers are thought to have lower quality lexical representations than skilled readers and lexical quality seems to be higher for high than for low frequency words. Figure 1 summarizes the main components of lexical representations described by the lexical quality hypothesis.

Figure 1. Key components of lexical quality hypothesis.



Note. Figure adapted from Perfetti & Hart (2002). The circles represent the main constituents of word representations. Phonological, orthographic and semantic word knowledge of the example word 'table' is shown in rectangular frames. Connections between circles represent constituent binding.

Investigations based on this framework have contributed to a better understanding of native language word knowledge in poor and skilled readers (e.g. Andrews, 2015; Murphy et al., 2016; Nelson & Perfetti, 2016; Richter et al., 2013). For example, Murphy et al. (2016) reported evidence supporting the idea that the quality of lexical representations varies across readers. They identified five profiles of lexical quality representations in English-speaking preschool students based on the measurement of orthographic, phonological, morphosyntactic and semantic information and used this variation to predict reading comprehension performance in Grade 1. Moreover, Goodwin, Gilbert, Cho and Kearns (2013) provided supporting evidence for the idea that the quality of lexical representations is item-specific and varies as a function of different psycholinguistic variables (e.g. higher lexical quality of high frequency vs. low frequency words; Perfetti & Hart, 2002; Perfetti, 2007). Goodwin et al. (2013) measured phonological, orthographic and semantic knowledge of English-speaking secondary students on 39 words. Item response modelling was used to be able to consider variability related to participants and words. Results revealed a significant contribution of reader (performance on language tasks) and word characteristics (frequency, orthographic-phonological and phonological opaqueness) to explain individual differences in word knowledge.

One limitation of past research embedded in the lexical quality framework is that the tasks used to measure different constituents of word representations involve more than one type of information rather than representing pure measures of phonological, orthographic or semantic knowledge. For example, Goodwin et al. (2013) used a word reading task to measure phonological word knowledge. However, word reading does not only require the involvement of the phonological constituent to access word pronunciations, but it also relies on the orthographic constituent to identify the written word forms. Similarly, word spelling, used to measure orthographic knowledge, involves the activation of the phonological constituent in addition to the orthographic constituent. This applies not only to Goodwin et al. (2013), but also to many other studies within the lexical quality framework (e.g. Murphy et al., 2016; Richter et al., 2013). Indeed, most tasks that have previously been used to assess word knowledge require the activation of more than one constituent within the lexical quality framework. These tasks, while not suitable to ask questions about the constituents themselves, provide insights into constituent binding. According to Perfetti (2017), constituent binding has so far not been explicitly investigated. He suggests that binding could simply be a consequence of the extent to which the three word constituents are specified or it could be related to additional cognitive mechanisms that exceed the quality of the components (p.64, Perfetti, 2017). While the concept of lexical quality seems very useful to reach a better understanding of the underlying cognitive mechanisms involved in word knowledge, Perfetti himself expresses that it "is a theoretical framework rather than a theory. Unlike a theory, LQ[6] does not lead directly to precise, testable predictions without additional assumptions" (p.53, Perfetti, 2017).

One way of sharpening the idea of constituent binding is to draw on language processing models. In contrast to the lexical quality framework, these models express precise and testable claims. This is especially well illustrated in the computational applications of these processing models, where assumptions have to be expressed as precise as possible to enable translation into programming algorithms (Caramazza & Coltheart, 2006). While these models also assume that lexical representations are built via the integration of phonological, orthographic and semantic information, they focus specifically on the processing steps that link different types of information to one another. This seems to be exactly what the lexical quality framework refers to under the term 'constituent binding'. For instance, in the language processing model by Ellis and Young (1988) spoken word comprehension is described as the activation of a phonological representation which is then mapped onto a semantic

[6] lexical quality

representation. In contrast, spoken word production consists of the activation of a semantic representation that is connected with a phonological representation. Word reading and spelling follow similar processing paths associating phonological and orthographic representations. Analyses based on this processing model have been widely used to explain performance dissociations in children with language difficulties, including children with poor literacy skills (Friedmann, Biran & Dotan, 2013; Friedmann & Coltheart, 2016; Kezilas, Kohnen, McKague & Castles, 2014; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003). This approach therefore seems to be useful to reach a better understanding of how different aspects of word knowledge are integrated and thus complement the concept of constituent binding.

Although the lexical quality framework has so far not been used to investigate foreign language word knowledge in children with poor literacy skills, existing evidence supports some of the predictions consistent with this framework. For example, van der Leij and Morfidi (2006) reported that only Dutch poor readers with poor English foreign language phonological and orthographic knowledge scored significantly below the control group on all foreign language tasks used in their study. In contrast, poor readers with poor foreign language phonological knowledge, but average orthographic knowledge were just as successful as the control group on several foreign language tasks (i.e. foreign language spoken word and sentence-picture matching, semantic fluency, serial rapid naming, speeded word and nonword reading, text reading accuracy and comprehension tasks). These findings lend support to the suggestion that the quality of lexical representation varies across readers, including in foreign language word knowledge in children with poor literacy skill. Nevertheless, available evidence on foreign language word knowledge in poor readers/spellers is predominantly based on group comparisons that average the performance of individual children with very different performances (e.g. Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Ho & Fong, 2005).

Another limitation in past research is that the tasks used in most studies have not been carefully matched to allow for comparisons across different aspects of foreign language word knowledge. For instance, different words have been used to measure different aspects of foreign language word knowledge (e.g. one set of words to measure spoken word comprehension and another set of words to measure written word comprehension). Furthermore, words used across tasks have not been matched on psycholinguistic variables (e.g. grammatical class, cognate status, syllable length) that have been shown to impact foreign language learning in unselected populations (de Groot & van Hell, 2005). This is problematic, because according to the lexical quality framework, lexical representations are item-specific and their quality varies as a function of different psycholinguistic variables (Perfetti & Hart, 2002; Perfetti, 2007, 2017). One way of addressing this issue is to use the same words across different tasks. Thus, different accuracy rates for the same word across tasks (e.g. correct spoken word production of (/teɪbl/), but incorrect spelling of the same word) cannot be a result of different word characteristics, but are more likely to indicate different underlying cognitive demands posed by each task.

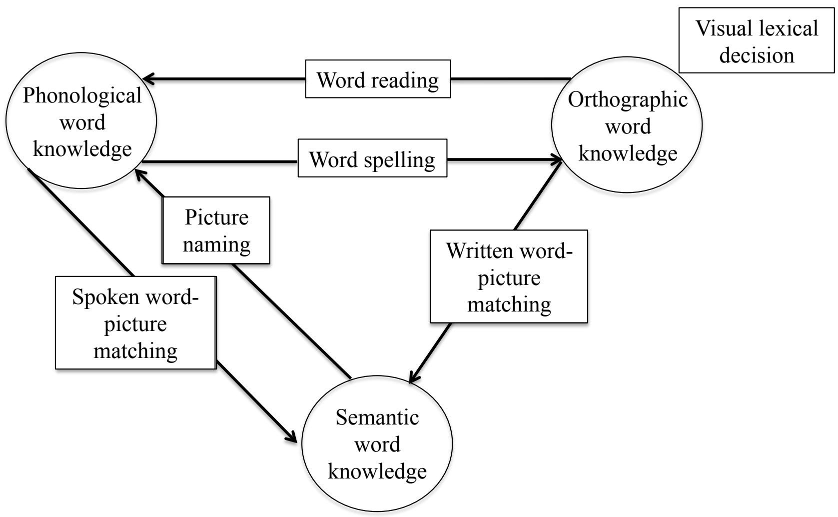
Furthermore, systematically varying the set of words across different psycholinguistic variables also makes it possible to assess the impact of these variables on foreign language word knowledge. While many word characteristics have been suggested to facilitate foreign language word learning (de Groot & van Hell, 2005), in this study we focus on the following five variables: (1) grammatical class, (2) cognate status, (3) presence of consonant clusters, (4) syllable length and (5) time-point of first exposure during foreign language instruction. To the best of our knowledge, the influence of these variables on foreign language word knowledge in children with poor literacy skills - to date - remains unknown.

In summary, available evidence on foreign language word knowledge of children with poor literacy skills provides limited information on a number of important questions, including (1) how variable the performance of individual children is across different tasks (e.g. comprehending a spoken word or spelling it), (2) in which foreign language tasks children show the strongest and the weakest performance and (3) if different word characteristics influence foreign language word knowledge of poor readers/spellers.

The present study

To address the above-mentioned questions, ten English speaking children with poor literacy skills who were learning German as a foreign language were asked to complete six different tasks involving the same 48 German words. In this study, we investigated some of the standard tasks administered in previous research, focusing on binding between phonological, orthographic and semantic word knowledge. Five of the tasks measured the binding of two of the components of word representations described by the lexical quality framework (Perfetti & Hart, 1992; Perfetti, 2007, 2017). As children with poor literacy skills are known to struggle with the acquisition of written word forms (Binamé, Danzio & Poncelet, 2015; Castles & Holmes, 1996; Suárez-Coalla, Ramos, Álvarez-Cañizo & Ramos, 2014; Wang, Marinus, Nickels & Castles, 2014), we decided to include one additional task to measure the orthographic knowledge constituent itself (see Figure 2).

Figure 2. Tasks used to measure lexical quality.



Performance patterns were analysed for each child individually and we addressed the following three research questions:

1. How variable is the lexical quality of foreign language word knowledge of poor readers/spellers? More specifically, how variable are children´s responses to the same words across six different tasks?

2. Which components of foreign language lexical representations show the strongest and the weakest bindings? To localize the sources of individual variability in the lexical quality of foreign language word representations, we identified the easiest and most difficult tasks at the group level and for each child. Based on this information, we identified individual strengths and weaknesses in the binding of different constituents of word representations.

3. Does the lexical quality of foreign language word knowledge in poor readers/spellers vary as a function of word characteristics? The set of 48 words was systematically varied according to grammatical class, cognate status, presence of consonant clusters, time-point of first exposure during foreign language instruction and syllable length. This allowed us to analyse the impact of word characteristics on the lexical quality of foreign language word representations.

Method

Participants

Participants were recruited from Grade 6 of a private primary school in Australia. In addition to regular English instruction, students received one German class of 45 minutes a week from the first year of formal schooling to Grade 6 (six complete years of German instruction). The same teacher taught all German classes. While this might have limited the extent to which our results can be applied to other foreign language instruction contexts, we can be sure that all children received relatively similar German instruction. This teacher was an English native-speaker with a highly proficient use of German. The fact that the teacher was not a native German speaker, as is typical for most foreign language instruction settings, may have influenced the children's perception and production of the German language. However, frequent exposure to authentic German language input was achieved through videos, audios and visits of German native-speakers. Although the focus of this study was not on foreign language instruction, we conducted a structured interview with the German teacher to obtain insights into children's exposure to German. Details of this interview are presented in the supplemental materials. The teacher reported that children were exposed to German in a playful way with a priority on developing communication skills. New vocabulary was predominantly introduced through pictures, avoiding translation into English. Opportunities for each student to speak German in every class were carefully planned through routine activities (e.g. saying good morning). Although the development of oral language skills was prioritized, children were regularly exposed to written German from Grade 3 of primary school onwards. The teacher explicitly taught new grapheme-phoneme correspondences (e.g. <z> = /ts/) that do not exist in English and frequently asked children to read and spell new words.

After gaining ethics approval from Macquarie University, we sent information letters, written consent forms and parent questionnaires to 63 parents of the three Grade 6 classes at the participating school. Only 14 parents agreed to their child´s participation. Exclusion criteria were uncorrected hearing or sight impairments and being bilingual with German as one of the native languages. None of the children who agreed to participate in the study showed any of these characteristics (as identified in a parent questionnaire). Inclusion criteria were below average scores (20^{th} percentile or below) on a standardized English native language reading or spelling test (see materials section below). Ten students met these inclusion criteria and were therefore categorized as poor readers/spellers. The remaining four students with average literacy skills were excluded. The unusually high percentage of participants with poor literacy skills is likely to be explained by a greater interest of parents of children with poor literacy skills to support research in this field, as opposed to parents of children with typical literacy skills.

Procedure

First, children completed a screening assessment measuring English native language reading and spelling skills. Based on these results, four children with typical literacy skills were excluded. The remaining ten children with poor literacy skills participated in further testing. We conducted a background assessment measuring English native language skills and broader cognitive skills, such as nonverbal reasoning and verbal short/working memory capacity. Children also completed six tasks measuring their foreign language word knowledge on a set of 48 words. Lastly, we collected additional information on children´s performance on foreign language nonword discrimination, repetition, reading and spelling tasks, as well as on their foreign language learning motivation. However, we do not report these additional foreign language measures here.

Testing took place at the beginning of the school year, during regular school hours, in a quiet room assigned by the school. Data for expressive native and foreign language skills was collected in one individual session and receptive native and foreign language skills were assessed in two group sessions. The first author, a native German speaker with a proficient knowledge of English, completed the assessment. All sessions lasted for 45 minutes each and the instructions were given in the language the children were being tested in. English native-language stimuli were presented by recordings from Australian native speakers.

Materials

Parent questionnaire. Parents were asked to complete a questionnaire with a multiple choice and brief answer format, on the child's developmental history, their linguistic background and the presence of any learning difficulties.

Native language screening measures.

Reading test. The *Castles and Coltheart 2 (CC2)* test (Castles, Coltheart, Larsen, Jones, Saunders & McArthur, 2009) was used to assess English native language reading skills. The test comprises 40 nonwords, 40 irregular words and 40 regular words. Items were presented in a random fixed order on separate printed cards (i.e. regular word, irregular word, nonword, irregular word, regular word, etc.). Children were asked to read each item aloud until they made five consecutive errors for one item type (e.g. five errors in a row on nonwords). The presentation of this item type was then concluded, while the other item types continued to be presented. Internal consistency is $\alpha = .85$, $\alpha = .85$ and $\alpha = .94$ for the regular, irregular and nonword subscales, respectively and validity is between $r = .67$ and $r = .69$ with the *Word Reading* and *Pseudoword Decoding* subtest of the *Wechsler Individual Achievement Test - Second Edition - Australian Standardised Edition* (WIAT-II; Harcount Assessment, 2007) as reported by Moore, Porter, Kohnen and Castles (2012). Children's responses were recorded and double scored by two Australian English native speakers. Disagreements were discussed between both raters to reach a consensus. As separate norms are available for each item type, in this study, poor readers were defined as children who performed at or below the 20th percentile on at least one item type (i.e. nonwords, irregular words, regular words).

Spelling test. The *Diagnostic Spelling Test irregular words (DiSTi)* (Kohnen, Colenbrander, Krajenbrink & Nickels, 2015) was used to assess English spelling ability. The test consists of a spelling to dictation task of 74 irregular words that have at least one letter with a spelling that does not follow the most common sound-letter mapping in English (for details see Kohnen et al., 2015). This test was administered in one group session and items were presented through a recording of a native Australian English speaker. Each word was first pronounced in isolation, then in a sentence and again in isolation. A stopping rule of five consecutive errors was applied during scoring. Internal consistency is reported as $\alpha = 0.94$, test-retest reliability as $r_s = .61$ and validity as $r = .61$ (Kohnen et al., 2015). Students who performed at or lower than the 20th percentile on this test were selected as poor spellers.

Background assessment. Participants were assessed with bespoke and standardized tasks on oral native language measures, nonverbal reasoning and verbal short/working memory capacity. A full description is provided in Table 1.

Table 1

Description of background assessment measures

Measure	Task	N of items	Description
Spoken word comprehension	Spoken word-picture matching	60	We used *Form A* of the *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)* by Dunn & Dunn (2007). Children heard a word recorded by an Australian English native speaker and were asked to mark the correct option between four pictures (phonological, semantic and unrelated distractor) on a response sheet. We adapted the task to a group administration format by projecting the pictures onto a screen. Therefore, all children completed the item sets 8 to 12, which cover the age ranges from 9 to 14 years. During scoring, the discontinuing rule of eight errors within an item set was applied. The internal consistency value reported for this test is $\alpha = .96$ and $\alpha = .98$ for the age range that we tested and test-re-test reliability is at $r = .91$ and $r = .92$. Despite having violated the individual administration conditions under which the normed values were constructed, we still decided to use the test norms as a reference instead of the raw scores. We used the British norms reported in the test manual.
Spoken word production	Picture naming	25	We used the subtest *Naming* from the *Assessment of Comprehension and Expression 6-11 (ACE 6-11)* (Adams, Cooke, Crutchley, Hesketh & Reeves, 2001). Children saw a picture and their response was elicited through a question (e.g. What is this?, Who is this?, What are these?). The internal consistency value reported for this subtest is reported as $\alpha = .78$ and test-retest reliability is $r = .84$ as reported in the test manual. Unfortunately no Australian norms are available and so we used the British norms reported in the test manual.
Nonverbal reasoning	Matrices	29	We used the subtest *Matrices* from the *Wechsler Intelligence Scale for Children - Fourth Edition Australian Standardised Edition (WISC IV Australian)* (Wechsler, 2005). In order to adapt this task to be administered in a group, we projected the items on a screen. Children saw an incomplete puzzle with five options for the missing piece and were asked to mark the correct option on the corresponding numbered box on a scoring sheet. To take into account the starting point indicated in the test manual for our participant group´s age range, we presented items number seven to 35 of the original task. Although children were asked to complete all items, the standard discontinuing rule indicated by the test manual was applied during the scoring procedure. Despite having violated the individual administration conditions under which the normed values were constructed, we still decided to use the test norms as a reference instead of the raw scores. This allowed for a more accurate measure of our participant´s performance, as the norm values take into account the age of each child. The internal consistency value for this test is $\alpha = .89$ to $\alpha = .92$ and test-retest reliability is reported as $r = .71$ to $r = .85$ for children of ten to twelve years of age
Verbal short/working memory	Forward and backward digit span	max. 32	This measure was assessed through the subtest *Forward and Backward Digit Span* from the *Wechsler Intelligence Scale for Children - Fourth Edition Australian Standardised Edition (WISC IV Australian)* (Wechsler, 2005). In the first part of this task, children were asked to repeat a chain of numbers that they were orally presented with. In the second part, they had to repeat a chain of numbers in the reverse order. The length of the number chain increased as the task progressed. Children heard the items from a recording on the computer to assure that the presentation was the same for all participants. The internal consistency value reported for this test is $\alpha = .86$ to $\alpha = .89$ and test-retest reliability is $r = .82$ to $r = .85$ for participants between 10 and 12 years of age.

Foreign language experimental measures. Participants completed six tasks on a set of 48 foreign language words. All items were selected from a list of vocabulary the children had learned and which was provided by the German teacher. Items were systematically varied according to: (1) grammatical class (16 nouns, 16 verbs, 16 adjectives), (2) cognate status (22 cognates and 26 non-cognates categorized according to a coding scheme by Kohnert, Windsor & Miller (2004) - see supplemental materials), (3) presence of consonant clusters (24 items with and without consonant clusters, respectively), (4) syllable length (16 monosyllabic and 32 bi- and tri-syllabic items) and (5) time-point of first exposure during foreign language instruction (24 early (before Grade 5) and 24 late instructed items (in Grade 5 and 6)). The complete list of words is presented in Appendices. First, picture naming and word reading were administered in an individual session. Then children completed the visual lexical decision and written word-picture matching task in the first group session. Finally, spoken word-picture matching and spelling were presented in the second group session. In order to minimise practice effects due to the repeated presentation of the same 48 words, we randomized the order of items between tasks and alternated the presentation of tasks with the background assessment measures. Responses to all tasks were coded as correct or incorrect.

Spoken word-picture matching. Children heard a German word from a recording spoken by a native-speaker and were asked to select a matching picture out of four options projected onto a screen. They registered their response on a response sheet representing the four options.

Picture naming. Participants saw a picture and heard a German sentence, naming the picture (*Das ist ein Tisch.* - translation: *This is a table*). They were then presented with a second picture and an incomplete sentence that was meant to elicit their picture naming response (*Das ist ein/e ......* - translation: *This is a/an ......*).

Written word-picture matching. Students saw four pictures and a written German word projected onto a screen. They were asked to select the correct picture matching the written word and mark their response on a response sheet.

Visual lexical decision. Children were presented with a written word projected onto a screen and had to mark on a response sheet if this was a real German word or a nonword. Ninety-six items were shown, half of them were real words, the other half were nonwords. Nonwords were created by systematically substituting the first vowel of each target word (e.g. *Hand - Hond, Feuer - Fauer, dünn - dönn*, etc.). Words and nonwords were presented in a randomized order. Each item was marked as correct only if both the word was accepted and its corresponding nonword rejected (e.g. *Hand* - response yes; *Hond* - response no).

Word reading. Participants were asked to read aloud the set of 48 words presented on a printed sheet of paper. Responses were recorded and judged by two German native speakers. The double scoring of 100% of the recordings revealed an average inter-rater-reliability of $\kappa = .65$.

Word spelling. Children heard the target words from a recording by a native German speaker. Each word was presented twice and participants were asked to write the word on an answer sheet.

Results

In this section, we first report descriptive statistics of the participants' background assessment and foreign language performance on the six tasks. Following this, we quantify the quality of item-specific foreign language lexical representations for each participant and identify strengths and weaknesses across tasks. Finally, we assess the impact of word characteristics on the quality of foreign language word knowledge.

Background assessment and foreign language tasks

Biodata of participants from the parent questionnaire, written and oral native language performance and nonverbal reasoning skills and verbal short/working memory capacity are presented in Tables 2 to 4.

Table 2

Biodata of participants obtained from parent questionnaire

Participant	Age	Gender	Linguistic background	Previous diagnosis
1	11 years 9 months	F	bilingual Dutch-English	n/a
2	11 years 7 months	F	monolingual	n/a
3	12 years 0 months	F	monolingual	dyslexia
4	11 years 0 months	M	monolingual	n/a
5	12 years 2 months	F	monolingual	n/a
6	12 years 0 months	F	monolingual	dyslexia
7	11 years 5 months	M	monolingual	dyslexia
8	10 years 10 months	F	monolingual	n/a
9	10 years 9 months	F	monolingual	n/a
10	11 years 6 months	F	monolingual	n/a

Note. F = Female; M = Male; n/a = not applicable.

Table 3

Descriptive statistics on native language literacy skills

Participant	Reading						Spelling	
	CC2 regular words		CC2 irregular words		CC2 nonwords		DiSTi irregular words	
	Raw score	*Pc*	Raw score	*Pc*	Raw score	*Pc*	Raw score	*Pc*
1	27	4	20	15	8	3	41	45
2	28	4	22	21	14	5	39	42
3	33	17	27	65	31	26	35	30
4	37	40	20	15	32	26	35	30
5	33	17	25	44	20	6	58	76
6	30	6	16	6	27	21	27	17
7	23	2	21	19	4	1	36	33
8	36	31	27	50	28	13	53	61
9	33	10	19	7	18	5	66	95
10	35	16	23	22	32	19	66	95

Note. Below average scores (at or below the 20th percentile) are marked in grey. CC2 = *Castles and Coltheart 2* test (2009); DiSTi = *Diagnostic Spelling Test irregular words* (Kohnen et al., 2015).

Table 4

Descriptive statistics on native oral language, nonverbal reasoning and verbal short/working memory capacity.

Participant	Spoken word-picture matching (n = 60)	Picture naming (n = 25)	Matrices (n = max. 29)	Digit span (n = max. 32)
1	-1.20	-1.33	0.00	-1.66
2	-1.00	1.00	-0.33	-0.66
3	-1.20	-1.00	-0.66	-1.00
4	-1.00	0.67	-0.66	-0.33
5	-1.40	-0.33	-0.33	-0.33
6	-1.40	0.00	-0.33	-1.00
7	-0.93	0.00	-1.33	-0.33
8	-1.00	1.00	-0.66	1.33
9	-1.27	1.33	-1.00	1.66
10	-0.87	1.00	-0.33	0.33

Note. Results are presented in z scores taking into account participant age according to norms reported in test manuals.

Based on the selection criteria for this study, all participants showed poor literacy skills (see Table 3). As would be expected, individual differences in their broader native language and cognitive profiles were also observed (see Table 4).

Quality of foreign language lexical representations

Unfortunately, one of the 48 words was accidentally omitted in one of the six experimental tasks. Therefore, analyses are based on the remaining 47 words. In Table 5 we present the participants' overall performance on these 47 words for each of the six tasks.

Table 5

Descriptive statistics on six tasks with 47 foreign language words for each participant.

Participant	Spoken word-picture matching	Picture naming	Written word-picture matching	Visual lexical decision	Word reading	Word spelling	*M*	*SD*
1	42	22	44	29	28	13	29.67	10.78
2	45	24	41	27	15	16	28.00	11.46
3	30	13	33	20	13	10	19.83	8.82
4	35	11	34	4	27	5	19.33	13.10
5	37	16	36	24	31	13	26.17	9.30
6	35	13	30	18	21	4	20.17	10.29
7	35	11	29	5	29	5	19.00	12.33
8	40	16	33	26	29	12	26.00	9.57
9	38	23	36	26	34	25	30.33	5.85
10	37	14	38	31	37	20	29.50	9.29
M	37.40	16.30	35.40	21.00	26.40	12.30		
SD	3.98	4.69	4.43	9.02	7.39	6.45		

Note. Results are presented in raw scores.

In order to quantify the quality of participants' lexical representations, we calculated the sum of correct responses for each item and participant. This value ranged from zero, indicating the lowest level of lexical quality equivalent to incorrect responses on all tasks, to six, the highest level of lexical quality resulting from correct responses on all six tasks. Results are reported separately for each participant and item in Table 6. Overall median, minimum and maximum values per participant and per item are also provided.

Table 6

Lexical quality across the six tasks per item and participant (ordered by item difficulty).

Item	Translation	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	*Mdn*	*Min*	*Max*
schön	beautiful	2	1	0	0	0	0	0	1	0	1	0.0	0	2
fressen	to eat	2	1	0	1	1	2	1	1	2	2	1.0	0	2
heiraten	to marry	2	0	0	0	1	1	0	1	4	1	1.0	0	4
sagen	to say	1	2	0	0	0	1	2	1	2	5	1.0	0	5
spielen	to play	4	1	0	1	0	2	0	1	2	4	1.0	0	4
wedeln	to wag	2	2	1	3	1	1	1	1	2	1	1.0	1	3
beissen	to bite	3	1	1	1	3	2	0	2	2	1	1.5	0	3
dünn	thin	2	5	2	0	2	0	1	0	3	1	1.5	0	5
gefährlich	dangerous	2	2	0	2	1	2	0	2	0	0	1.5	0	2
Dreieck	triangle	2	2	0	3	3	1	1	2	4	3	2.0	0	4
Feuer	fire	2	2	0	3	1	1	2	2	3	2	2.0	0	3
klettern	to climb	2	3	1	1	2	3	1	3	1	4	2.0	1	4
lockig	curly	3	2	2	1	3	1	2	2	2	5	2.0	1	5
mutig	brave	1	2	2	2	2	3	3	3	1	1	2.0	1	3
schlafen	to sleep	3	2	1	2	3	1	2	2	3	3	2.0	1	3
Viereck	square	2	2	2	2	1	1	2	1	3	4	2.0	1	4
wild	wild	3	2	3	2	3	2	1	2	2	3	2.0	1	3
heissen	to be called	2	3	1	0	3	1	1	4	3	3	2.5	0	4
Kreis	circle	5	4	0	3	1	2	2	3	5	2	2.5	0	5
Bauch	belly	5	3	3	1	2	2	2	3	4	3	3.0	1	5
Drachen	dragon	4	3	2	2	2	2	4	3	5	5	3.0	2	5
essen	to eat	5	3	2	2	4	0	2	4	3	5	3.0	0	5
flattern	to flap	3	2	3	3	2	3	2	1	4	3	3.0	1	4
freundlich	friendly	2	6	2	4	5	2	2	3	4	3	3.0	2	6
mittelgross	medium tall	3	4	2	3	3	1	3	4	4	4	3.0	1	4
tanzen	to dance	6	3	2	2	4	3	2	1	4	3	3.0	1	6
Bein	leg	4	6	0	3	3	0	5	5	4	3	3.5	0	6
springen	to jump	5	4	3	3	3	4	3	4	4	3	3.5	3	5
trinken	to drink	4	4	2	1	3	3	2	4	6	4	3.5	1	6
fantastisch	fantastic	4	4	3	4	5	3	2	4	5	4	4.0	2	5
Hand	hand	4	6	4	3	5	4	2	3	4	4	4.0	2	6
Hexe	witch	3	3	4	5	5	2	3	5	5	4	4.0	2	5
Schwein	pig	4	4	3	3	6	3	4	5	6	6	4.0	3	6
singen	to sing	6	4	4	4	5	5	3	4	5	6	4.5	3	6
Affe	monkey	5	5	3	3	6	4	3	6	5	6	5.0	3	6
Elefant	elephant	6	5	5	3	6	5	4	4	6	6	5.0	3	6
Gelb	yellow	6	6	6	3	4	3	4	6	6	4	5.0	3	6
Kuh	cow	5	5	3	3	5	3	4	6	6	5	5.0	3	6
Lang	long	5	6	5	6	4	6	3	4	5	6	5.0	3	6
Prima	great	6	6	5	3	5	3	3	4	6	5	5.0	3	6
schwimmen	to swim	5	4	5	4	5	5	4	5	6	5	5.0	4	6
Wolf	wolf	3	6	4	4	5	5	5	6	4	6	5.0	3	6
Blau	blue	6	5	5	3	6	5	5	6	6	6	5.5	3	6
orange	orange	6	5	5	4	6	3	5	6	6	6	5.5	3	6
Ratte	rat	6	5	6	3	6	6	3	5	6	5	5.5	3	6
rot	red	6	6	6	4	5	5	5	6	4	6	5.5	4	6
Schlange	snake	6	6	6	3	6	4	3	5	6	5	5.5	3	6
Mdn		4.0	4.0	2.0	3.0	3.0	2.0	2.0	3.0	4.0	4.0			
Min		1	0	0	0	0	0	0	0	0	0			
Max		6	6	6	6	6	6	5	6	6	6			

Note. Items are presented in increasing order according to the median lexical quality value for all participants.

Binding of components within foreign language lexical representations

The second research question aimed to identify stronger and weaker bindings between the components of foreign language lexical representations in poor readers/spellers. Although our main interest was to investigate constituent binding, we also included a measure of children´s orthographic knowledge in these analyses, as based on our inclusion criteria participants were expected to show difficulties with written word forms. For this purpose, we analysed group and individual data to identify which tasks children struggled with the most and the least. For the group analyses, we used linear mixed-effects modelling in R (Baayen, 2008), as implemented in lme4 package (Bates, Maechler, Bolker, & Walker, 2014). The fixed effect was task, with the six different tasks as levels of this independent variable (i.e., reading, spelling, spoken word-picture matching, picture naming, written word-picture matching, visual lexical decision). The dependent variable was item-level accuracy (correct or incorrect). We fit logistic regression models. As we were interested in taking into account item- and participant-specific variability, we entered participants and items as random slopes into the model. The analysis was based on 2820 observations from 10 children and 47 items. Finally, we ran post-hoc contrasts to determine significant differences between children's performance on the six tasks. Results are shown in Table 7 and depicted in Figure 3.

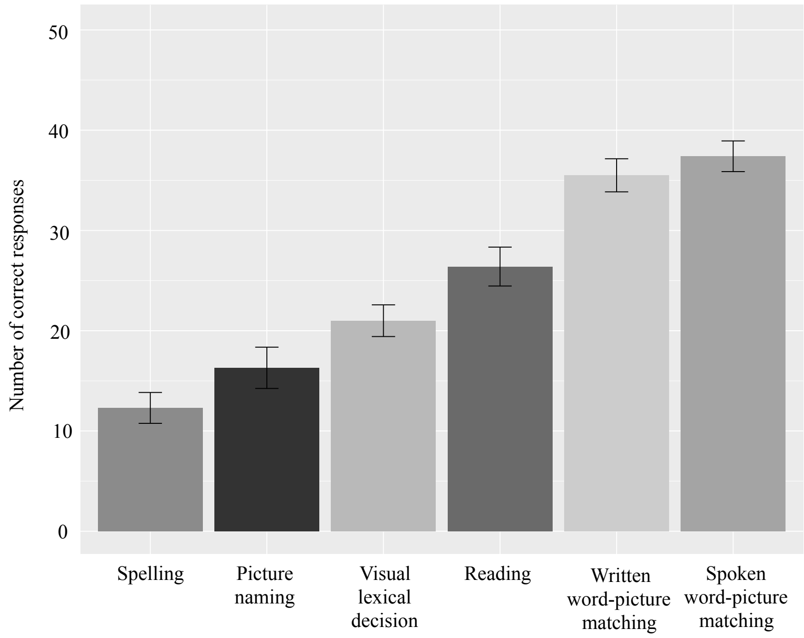
Table 7

Results of linear mixed model on foreign language word knowledge with random slopes for participant and item data with post-hoc contrast analyses.

Task	I.	II.	III.	IV.	V.	VI.
I. Written word-picture matching	-					
Estimate	-					
SE	-					
z	-					
p	-					
II. Word spelling						
Estimate	3.7911					
SE	0.4088					
z	9.274					
p	<.0010***					
III. Spoken word-picture matching						
Estimate	-0.4106	-4.2018				
SE	0.3411	0.4768				
z	-1.204	-8.813				
p	.3622	<.0010***				
IV. Word reading						
Estimate	1.6162	-2.1749	2.0268			
SE	0.5357	0.4760	0.5683			
z	3.017	-4.569	3.567			
p	.0127*	<.0010***	.0025**			
V. Visual lexical decision						
Estimate	1.6162	-1.3537	2.8481	0.8213		
SE	0.5357	0.3019	0.4681	0.5293		
z	3.017	-4.484	6.084	1.552		
p	.0127*	<.0010***	<.0010***	.3622		
VI. Picture naming						
Estimate	3.2852	-0.5059	3.6959	1.6690	0.8478	-
SE	0.3475	0.3728	0.4071	0.5329	0.3944	-
z	9.454	-1.357	9.079	3.132	2.150	-
p	<.0010***	.3622	<.0010***	.0104*	.1263	-

Note. Adjusted p values are reported with *holm* method correction for multiple comparisons. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Figure 3. Comparisons between performances across six tasks on 47 foreign language words



Note. Error bars represent 95% confidence intervals.

The lowest performance was identified for word spelling, followed by picture naming, spoken word-picture matching, visual lexical decision, word reading, written word-picture matching and finally spoken word-picture matching. Although the participant coefficients from the above-mentioned linear mixed modelling analysis roughly showed the same patterns of strengths and weaknesses for each child, for some participants the order of task difficulty differed from the group results (see participants 1 and 2 in Table 8).

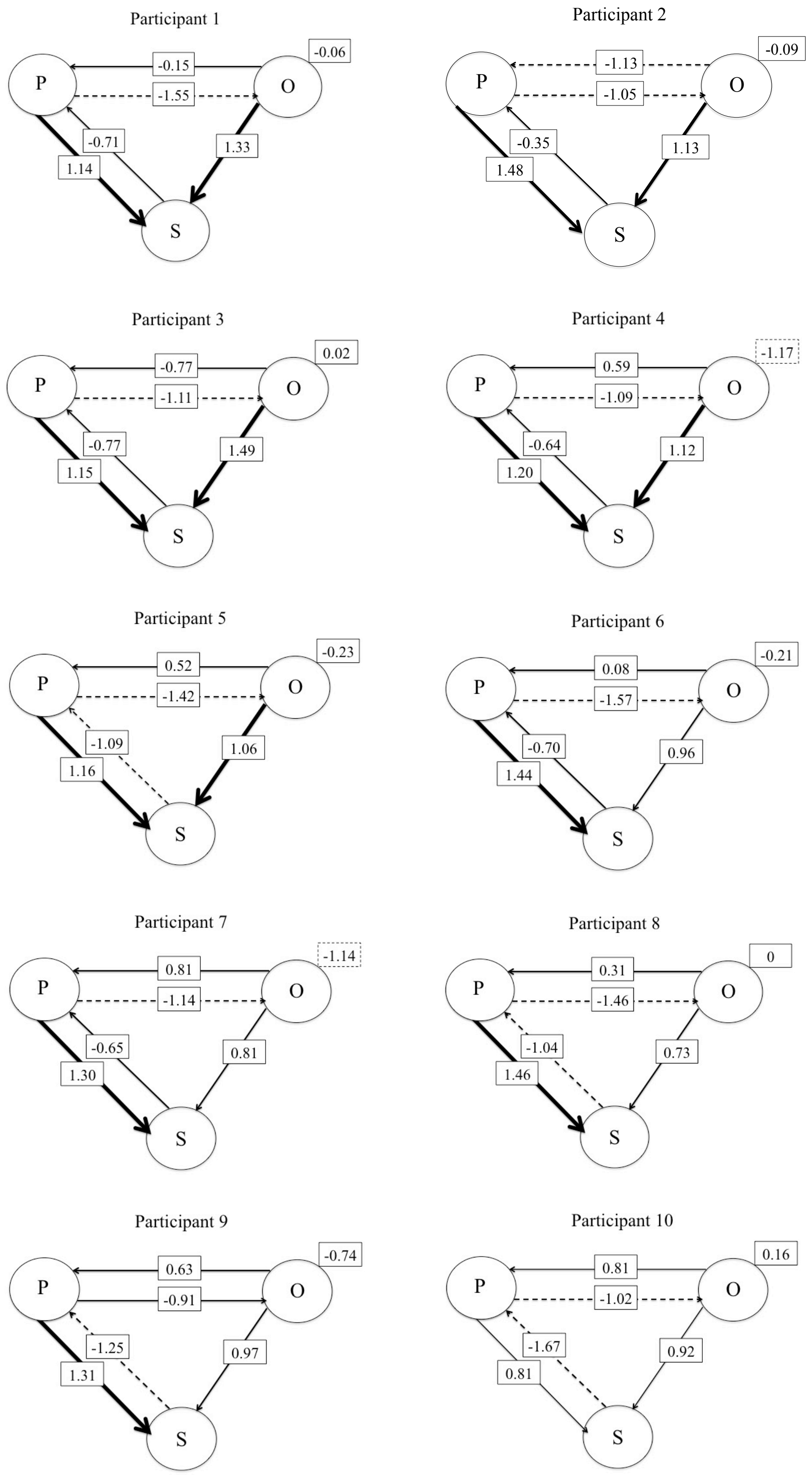
Table 8

Participant coefficients of linear mixed model results on foreign language word knowledge.

Participant	Written word-picture matching	Word spelling	Spoken word-picture matching	Word reading	Visual lexical decision	Picture naming
1	3.38	-0.77	3.92	0.84	0.88	-0.31
2	3.21	-1.18	3.67	-0.96	0.49	0.04
3	3.04	-0.47	3.17	0.47	1.26	-1.64
4	3.37	-1.01	4.01	2.48	-0.35	-2.01
5	3.28	-0.12	3.61	2.21	1.25	-1.23
6	3.31	-0.78	3.77	1.59	1.10	-1.91
7	3.37	-0.80	3.96	2.94	-0.07	-2.22
8	3.31	-0.28	3.69	1.94	1.38	-1.22
9	3.12	0.40	3.22	2.01	1.04	-0.52
10	3.26	0.53	3.44	2.97	1.86	-1.08

Therefore, we also computed analyses for each participant separately. For this purpose, scores that fell above and below one standard deviation of each participant's mean performance across all six tasks were categorized as intra-individual strengths and weaknesses, respectively (see Table 5 for *M* and *SD* for each participant). Results are shown in Figure 4 with different arrows representing strong, weak and average-strength bindings between the components of foreign language lexical representations. The values added to each arrow show the difference (measured in standard deviations) between the performances in each task from the mean performance across all six tasks.

Figure 4. Strength of binding across constituents of foreign language lexical representations.



Note. P = Phonological word knowledge; O = Orthographic word knowledge; S = Sematic word knowledge; Dashed arrows represent weak bindings; thick arrows represent strong bindings; medium thick arrows represent average bindings. Values for each arrow express the extent to which the performance on each task differed from the mean performance across all six tasks, as measured in standard deviations.

The following four types of intra-individual weaknesses were found: (a) binding from phonological to orthographic word knowledge as measured by the word spelling task (9/10 participants - see Figure 4 all participants except participant 9); (b) binding from semantic to phonological word knowledge as measured by the picture naming task (4/10 participants - see Figure 4 participants 5, 8, 9 and 10); (c) orthographic word knowledge as measured by the visual lexical decision task (2/10 participants - see Figure 4 participants 4 and 7) and (d) binding from orthographic to phonological word knowledge as measured by the word reading task (1/10 participants - see Figure 4 participant 2). In contrast, intra-individual strengths were identified for (a) binding from phonological to semantic word knowledge as measured by the spoken word-picture matching task (9/10 participants - see Figure 4 all participants except participant 10) and (b) binding from orthographic to semantic word knowledge as measured by the written word picture matching task (5/10 participants - see Figure 4 participants 1, 2, 3, 4 and 5).

Impact of word characteristics on foreign language word knowledge

The third research question of this study asked if the quality of foreign language word knowledge in poor readers/spellers varied as a function of word characteristics. To answer this question, we computed separate Chi-square and Fisher's exact tests with the five item characteristics that were systematically varied across the 47 target words as independent variables: (1) grammatical class, (2) cognate status, (3) presence of consonant clusters, (4) syllable length and (5) time-point of first exposure during foreign language instruction. The number of correct and incorrect responses across the six tasks achieved by each participant was the dependent variable for these analyses. Moreover, a Bonferroni correction for multiple comparisons was applied to each participant's results. Results are presented in Table 9 and depicted in Figure 5.

Table 9

Results from Chi-square and Fisher's exact test on the influence of word characteristics on response accuracy across the six experimental tasks.

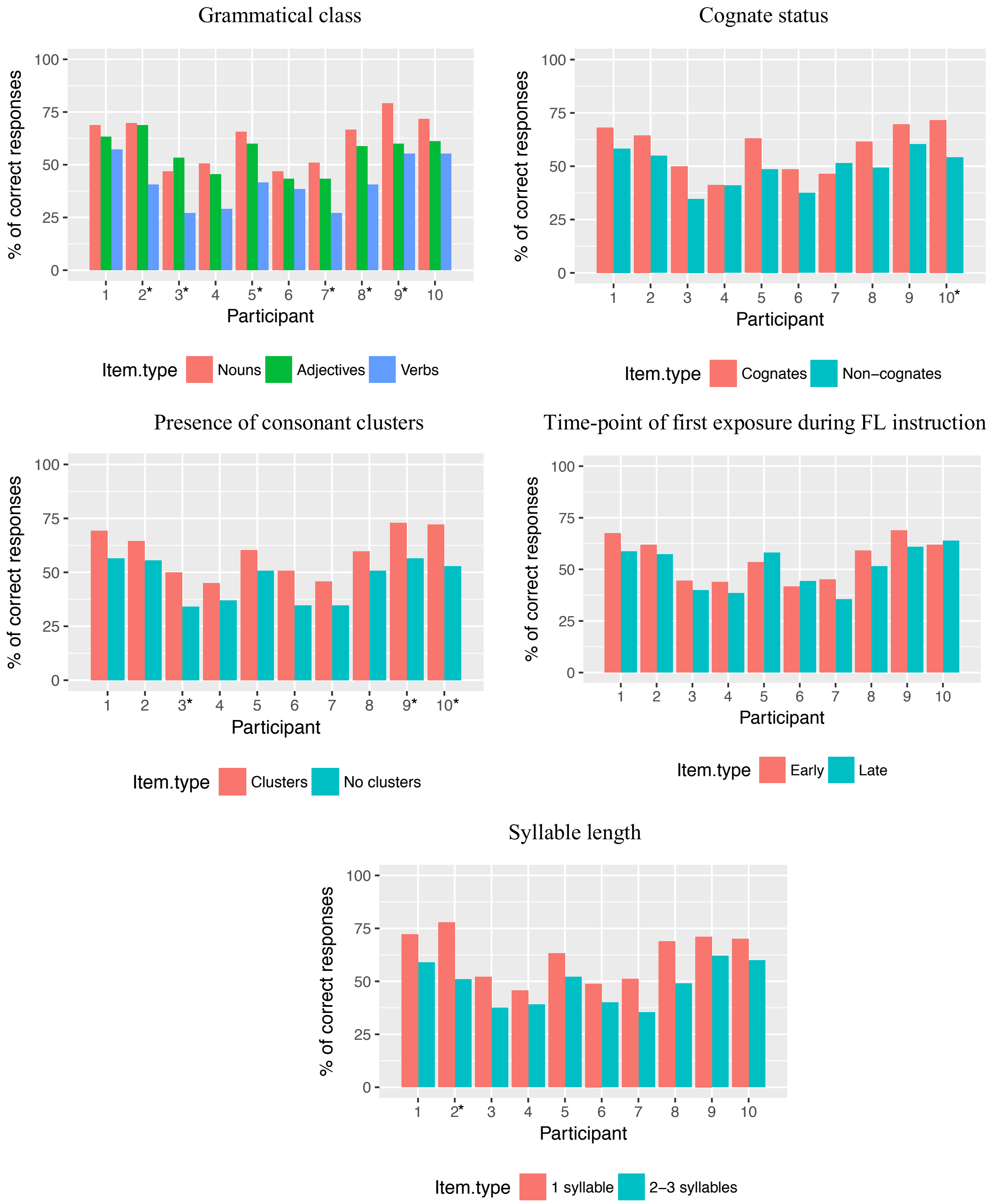
Participants	Grammatical class						Cognate status			
	% of correct responses			$\chi 2$	*p*	Follow-up Fisher's exact test	% of correct responses		*z*	*p*
	Nouns (*n* = 16)	Adjectives (*n* = 15)	Verbs (*n* = 16)				Cognates (*n* = 23)	Non-cognates (*n* = 24)		
1	68.7	63.3	57.2	2.7	.257		68.1	58.3	1.5	.115
2	69.7	68.8	40.6	21.7	<.001*	N > V**; A > V**	64.5	54.9	1.5	.127
3	46.8	53.3	27.0	14.4	<.001*	A > V**	50.0	34.7	2.4	.013
4	50.5	45.5	29.1	8.8	.012		41.3	41.0	0.1	1.000
5	65.6	60.0	41.6	12.1	.002*	N > V**	63.0	48.6	2.3	.020
6	46.8	43.3	38.5	1.3	.503		48.6	37.5	1.7	.080
7	51.0	43.3	27.0	11.9	.002*	N > V**	46.4	51.4	0.7	.470
8	66.6	58.8	40.6	13.8	<.001*	N > V**	61.6	49.3	1.9	.051
9	79.1	60.0	55.2	13.4	.001*	N > V**	69.6	60.4	1.4	.138
10	71.8	61.1	55.2	5.8	.053		71.7	54.2	2.9	.003*

Participants	Presence of consonant clusters				Time-point of first exposure during FL instruction				Length			
	% of correct responses		*z*	*p*	% of correct responses		*z*	*p*	% of correct responses		*z*	*p*
	Clusters (*n* = 23)	No clusters (*n* = 24)			Early (*n* = 24)	Late (*n* = 23)			1 syllable (*n* = 15)	2-3 syllables (*n* = 32)		
1	69.4	56.5	2.1	.033	67.4	58.7	1.3	.167	72.2	58.9	2.0	.042
2	64.6	55.6	1.4	.157	61.8	57.2	0.6	.511	77.8	51.0	4.1	<.001*
3	50.0	34.1	2.5	.009*	44.4	39.9	0.6	.510	52.2	37.5	2.2	.027
4	45.1	37.0	1.2	.203	43.8	38.4	0.7	.430	45.6	39.1	0.9	.367
5	60.4	50.7	1.5	.129	53.5	58.0	0.6	.522	63.3	52.1	1.6	.100
6	50.7	34.8	2.5	.010	41.7	44.2	0.3	.757	48.9	40.1	1.2	.208
7	45.8	34.8	1.7	.077	45.1	35.5	1.5	.127	51.1	35.4	2.3	.017
8	59.7	50.7	1.3	.162	59.0	51.4	1.1	.247	68.9	49.0	3.0	.002
9	72.9	56.5	2.7	.005*	68.8	60.9	1.2	.208	71.1	62.0	1.3	.173
10	72.2	52.9	3.2	.001*	61.8	63.8	0.2	.828	70.0	59.9	1.5	.132

Note. *p < .01 and **p <.007 Bonferroni corrected *p*-value controlling for five and seven comparisons per participant, respectively.

N = Nouns; A = Adjectives; V = Verbs

Figure 5. Influence of word characteristics on response accuracy across the six experimental tasks.



Note. *p < .01 Bonferroni corrected *p*-value controlling for five comparisons per participant; FL = Foreign language.

Six of ten participants showed a significant impact of grammatical class on foreign language word knowledge. This effect was driven by a lower performance for verbs than nouns for five children (see participants 2, 5, 7, 8 and 9 in Table 9) and by lower scores on verbs than adjectives for two children (see participant 2 and 3 in Table 9). Although we found no evidence of a grammatical class effect for the remaining four children, they showed a similar trend with a lower proportion for verbs than adjectives and nouns (see Figure 5). With regard to cognate status, only participant 10 was better at processing cognates than non-cognates (see Table 9). Eight of the nine remaining participants showed the same trend and one participant scored higher on non-cognate items as compared to cognate items (see Participant 7 in Figure 5). Three participants reached significantly higher accuracy rates for words with consonant clusters than words without consonant clusters (see participant 3, 9 and 10 in Table 9). The same trend was observed for the remaining seven participants (see Figure 5). No significant impact of time-point of first exposure during foreign language instruction on children's foreign language word knowledge was found. Seven of ten participants revealed a trend towards higher scores on words that were introduced at an earlier, as compared to a later time-point (see participants 1, 2, 3, 4, 7, 8, 9 in Table 9). In contrast, the remaining three participants showed the opposite pattern (see participants 5, 6 and 10 in Table 9). Finally, concerning the influence of syllable length on children's foreign language word knowledge, we observed a significant advantage of monosyllabic as compared to bi- and tri-syllabic words only for participant 2 (see Table 9), although a trend towards the same pattern was observed for all participants (see Figure 5).

Discussion

The present study investigated the quality of foreign language word knowledge in children with poor reading and spelling skills. More specifically, we asked (1) how variable the quality of foreign language word knowledge of individual poor readers/spellers is across different tasks (e.g. comprehending a spoken word or spelling it), (2) which constituent bindings represent individual strengths and weaknesses in foreign language lexical representations of children with poor literacy skills and (3) if different word characteristics influence foreign language word knowledge of poor readers/spellers? In the following sections we discuss our findings in turn.

First, we found the quality of foreign language word knowledge of children with poor literacy skills to vary across individuals and words. This was expected, as poor readers/spellers have been described as a heterogeneous group comprising of children with different underlying cognitive deficits (e.g. Friedmann & Coltheart, 2016; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; McArthur et al., 2013; Sotiropoulos & Hanley, 2017). Variability across words was also expected based on the assumption of the lexical quality framework that word representations are item-specific. Although these findings are not surprising from the perspective of the lexical quality framework, they underline the need to take into account individual and item variability when analysing foreign language word knowledge in children with poor literacy skills. This is important, because previous findings are mostly based on group averages that merge the responses of different poor readers/spellers to different sets of foreign language words across tasks (e.g. Farukh & Vulchanova, 2016; Ghonsooly & Javadian, 2010; Ho & Fong, 2005). They may therefore be reflecting an inaccurate picture of foreign language word knowledge in poor readers/spellers, which would also have consequences in terms of practical implications (von Hagen et al., in preparation; also see Andrews, 2015). Methodological approaches that inherently take into account individual variability (i.e. linear mixed modelling analyses, item response modelling, single case or case series studies) might be in a better position to capture individual differences in foreign language word knowledge in poor readers/spellers.

Regarding our second research question, results from linear mixed modelling analyses (taking into account individual and item variability) indicated two general trends that were also identified in the individual profiles of nine of the ten participants. First, binding from phonological to orthographic word knowledge (measured by the word spelling task) was found to be the weakest constituent binding. In terms of language processing models, such as

Ellis and Young's (1988) model, these results could be reflecting underlying difficulties either in the pathway that connects phonological and orthographic information or in the orthographic and phonological knowledge constituents themselves. This is consistent with previous work that suggests underspecified phonological or orthographic representations or difficulties in accessing these representations as underlying deficits in children with poor literacy skills (Friedmann & Lukov, 2016; Kohnen et al., 2018; McCloskey & Rapp, 2017; Zoccolotti & Friedmann, 2010).

For two children difficulties seemed to be located at the level of the orthographic constituent itself, as evidenced in intra-individual weaknesses in the visual lexical decision task and word spelling task (see participant 4 and 7 in Figure 4). For another child (see participant 2 in Figure 4) the source of low quality foreign language word knowledge seemed to be located at the connections between phonological and orthographic information, as evidenced in intra-individual weaknesses in the word reading and spelling tasks. While it could be possible that deficits in orthographic word representations were the source of difficulties in word reading and spelling, the child’s relatively good performance on the visual lexical decision task rules this out. In a similar way, an underlying deficit at the level of the phonological constituent itself seemed unlikely, as the child achieved a relatively good performance on the picture naming and spoken word-picture naming tasks that both rely on phonological knowledge. The source of the child´s difficulties in word reading and spelling therefore seemed to be located at the level of constituent binding between phonological and orthographic knowledge. In coherence with the lexical quality framework, it would therefore be likely this child to have difficulties in linking phonological and orthographic word forms. However, within Ellis and Young's (1988) model another possibility exists. Difficulties in word reading and spelling can also be caused by difficulties in grapheme-phoneme and phoneme-grapheme conversion mechanisms. German words predominantly consist of regular grapheme-phoneme correspondences and can therefore be successfully read and spelled by grapheme-phoneme/phoneme-grapheme conversion without activating item-specific phonological or orthographic word representations (Landerl, 2017). This alternative explanation based on Ellis and Young´s (1988) is another reason to complement the lexical quality framework with contributions from language processing models.

The second general trend revealed by our group analyses was that binding from phonological to semantic word knowledge (measured by the spoken word-picture matching task) was the strongest constituent binding. This was consistent at an individual level for nine of ten participants (see all except participant 10 in Figure 4). Five of ten participants also showed intra-individual strengths in binding from orthographic to semantic word knowledge (measured by the written word-picture matching task) (see participants 1 to 5 in Figure 4). It is possible that for participant 1, the only child with a bilingual background in this study, previous experience with Dutch, a language that shares many cognates with German, played a role in facilitating binding from phonological and orthographic to semantic word knowledge. In contrast, binding from semantic to phonological information (measured by the picture naming task) emerged as an intra-individual weakness for four of ten participants (see participants 5, 8, 9 and 10 in Figure 4).

This is in line with extensive research reporting an advantage of receptive over expressive language skills (e.g. Monsell, 1987; Roelofs, 2003; Shallice, McLeod & Lewis, 1985). While imprecise phonological information could be enough to successfully comprehend an orally presented foreign language word, more detailed phonological knowledge seems to be necessary to produce the same spoken word. Underlying difficulties in the phonological knowledge constituent itself or in the pathway that connects semantic and phonological information might therefore explain contrasts between receptive and expressive foreign language word knowledge.

Our results could, at least in part, be a consequence to task difficulty. While the spoken and written word-picture matching and the visual lexical decision tasks required children to choose the correct response between different options, the picture naming, word reading and spelling tasks imposed higher demands as oral or written responses had to be produced. However, this explanation cannot explain the results entirely as seven of ten participants achieved higher scores on some of the harder tasks (word reading) than on the easier ones (visual lexical decision) (see participants 4 to 10 in Figure 4).

With respect to the third research question, we found a significant impact of grammatical class, cognate status, presence of consonant clusters and syllable length, but not time-point of first exposure during foreign language instruction on children's overall foreign language word knowledge. Although these effects only reached significance for some participants, general trends were consistent for most children (see Table 9 and Figure 5). Similar to previous studies on foreign language word learning in unselected populations (i.e., not specifically looking at poor reader/spellers), we found higher accuracy for (a) nouns and adjectives, as compared to verbs, (b) cognates as compared to non-cognates and (c) monosyllables as compared to bi- and tri-syllables (de Groot & van Hell, 2005).

A surprising finding was that children were more accurate on words with as opposed to words without consonant clusters. As consonant clusters add phonological complexity to a word and many children with poor literacy skills show phonological deficits, one may have expected words with consonant clusters to be more difficult. This finding should be interpreted with caution, however, as our analyses were based on children´s overall performance across tasks. It is likely that the presence of consonant clusters makes tasks that include the phonological constituent (e.g. picture naming, word spelling) harder, but plays less of a role in other tasks (e.g. visual lexical decision). In the context of the current study, it was not possible to conduct further analyses to disentangle this issue, as our design focused on analysing individual performance patterns of poor readers/spellers' foreign language word knowledge and therefore data was not sufficient to conduct larger group based analyses.

However, item response modelling seems to be a promising way forward to obtain a more detailed picture in this respect in the future (see for example Goodwin et al., 2013).

Lastly, we found no influence of time-point of first exposure during foreign language instruction on children's overall foreign language word knowledge. This variable was included as a proxy for age of acquisition, which has been shown to be an important predictor of vocabulary development (see Juhasz, 2005 for a review). However, it might have rather been a measure of age of exposure than acquisition, as the classification into "early" and "late" words was based on a list provided by the German teacher with information of the sequence of teaching those words. Also, “early” and “late” exposure may be a too crude distinction to capture the effects of age of exposure. Future studies collecting longitudinal data might be in a better position to measure the impact of age of acquisition and/or exposure on foreign language word knowledge in children with poor literacy skills.

To summarize, the present study provides new insights into foreign language word knowledge in children with poor literacy skills. Our findings show that foreign language word knowledge varies across poor readers/spellers and words. Underlying difficulties of different components of foreign language word processing may explain individual differences across tasks. Furthermore, word characteristics that have been reported to influence foreign language word learning in unselected populations play a similar role in children with poor literacy skills.

From a theoretical perspective, the present study adds to existing work on the lexical quality framework by underlining that lexical representations are reader- and item-specific. This highlights the need for future studies to consider individual and item variability when analysing native and foreign language word knowledge. Moreover, this study extends the lexical quality framework to foreign language word knowledge in children with poor literacy skills. We also propose that the concept of constituent binding proposed by Perfetti and Hart (2002) can be sharpened by defining the tasks within models of language processing, such as the one by Ellis and Young (1988). For example, this could guide future studies that may wish to include more specific testing of all of the components and their possible bindings, expanding on the current study.

From a practical perspective, our results indicate that there are many reasons why children with poor literacy skills can have difficulties in successfully using foreign language words across different tasks. An identification of the level of breakdown during foreign language word processing guided by language processing models can provide a better understanding of the tasks that might be easier or more difficult for individual children. Furthermore, similarly to children with typical literacy skills, poor readers/spellers as a group seem to be more successful with nouns and adjectives, as opposed to verbs, with cognates, as opposed to non-cognates and with monosyllables, as opposed to bi- and tri-syllables. This information can aid foreign language teachers in the implementation of support strategies for children with poor literacy skills.

References

Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of comprehension and expression 6-11.* London: GL assessment.

Andrews, S. (2015). Individual Differences Among Skilled Readers: The Role of Lexical Quality. In Alexander Pollatsek, Rebecca Treiman (Eds.), *The Oxford Handbook of Reading*, (pp. 129-148). Oxford: Oxford University Press

Baayen, R. H. (2008). *Analysing linguistic data: A practical introduction to statistics using R.* Cambridge: University Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4.* R package version, 1(7), 1-23. Retrieved from http://CRAN.R-project.org/package=lme4.b.

Binamé, F., Danzio, S. & Poncelet, M. (2015) Relative Ease in Creating Detailed Orthographic Representations Contrasted with Severe Difficulties to Maintain Them in Long-term Memory Among Dyslexic Children. *Dyslexia*, 21, 361–370. doi: 10.1002/dys.1506.

Caramazza, A. & Coltheart, M. (2006) Cognitive Neuropsychology twenty years on. *Cognitive Neuropsychology*, 23(1), 3-12. doi: 10.1080/02643290500443250

Castles, A., Coltheart, M., Larsen, L., Jones, P., Saunders, S., & McArthur, G. (2009). Assessing the basic components of reading: A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties*, *14*(1), 67-88. doi: 10.1080/19404150902783435

Castles, A., & Holmes, V. M.(1996). Subtypes of developmental dyslexia and lexical acquisition. *Australian Journal of Psychology*, *48*, 130–135. doi: 10.1080/00049539608259519

De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in second language acquisition*, 309-329.

de Groot, A. M. B., & van Hell, J. G. (2005). The Learning of Foreign Language Vocabulary. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9-29). New York, NY, US: Oxford University Press.

Dunn, L.M., & Dunn, D.M. (2007). *The Peabody Picture Vocabulary Test, Fourth Edition.* Bloomington, MN: NCS Pearson, Inc.

Ellis, A.W. & Young, A.W. (1988) *Human cognitive neuropsychology*. Hove: Erlbaum.

Farukh, A., & Vulchanova, M. (2016). L1, quantity of exposure to L2 and reading disability as factors in L2 oral comprehension and production skills. *Learning and Individual Differences*, *50*, 221-233. doi: 10.1016%2Fj.lindif.2016.07.013

Friedmann, N., & Coltheart, M. (2016). Types of developmental dyslexia. In A. Baron, & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives*. Berlin, Boston: De Gruyter Mouton.

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, 44(9), 1146-1160. doi: 10.1016/j.cortex.2007.09.005

Friedmann, N., Biran, M. & Dotan, D. (2013). Lexical retrieval and breakdown in aphasia and developmental language impairment. In C. Boeckx & K.K. Grohmann (Eds.), *The Cambridge handbook of biolinguistics* (pp.350-374). Cambridge University Press.

Ghonsooly, B. & Javadian, M. (2010). An Examination of Developmental Dyslexia among Iranian EFL Second Graders. *International Journal of Language Studies*, 4(4*),* 333-356.

Goodwin, A. P., Gilbert, J. K., Cho, S. J., & Kearns, D. M. (2014). Probing lexical representations: Simultaneous modeling of word and reader contributions to multidimensional lexical representations. *Journal of Educational Psychology*, *106*(2), 448. doi: 10.1037/a0034754

Harcourt Assessment. (2007). *The Wechsler Individual Achievement Test – Second Edition – Australian Standardised Edition*. Sydney, Australia: Author.

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

Ho, C. S. H., & Fong, K. M. (2005). Do Chinese dyslexic children have difficulties learning English as a second language? *Journal of Psycholinguistic Research*, 34(6), 603-618. doi: 10.1177/0022219409345018.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological bulletin*, *131*(5), 684-712. doi: 10.1037/0033-2909.131.5.684

Kezilas, Y., Kohnen, S., McKague, M., & Castles, A. (2014). The locus of impairment in English developmental letter position dyslexia. *Frontiers in human neuroscience*, *8*, 356. doi: 10.3389/fnhum.2014.00356

Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in Grades 1-7. *Australian Journal of Learning Difficulties*, 20(1), 15-38. doi:10.1080/19404158.2015.1023209

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime' becomes 'smile': Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681–3692. doi:10.1016/j.neuropsychologia.2012.07.016

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex*. doi: 10.1016/j.cortex.2018.04.008

Kohnert, K., Windsor, J., & Miller, R. (2004). Crossing borders: Recognition of Spanish words by English-speaking children with and without language impairment. *Applied Psycholinguistics*, *25*(4), 543-564. doi: 10.1017/S0142716404001262

König, E. & Gast, V. (2012). *Understanding English-German Contrasts*. Berlin: Erich Schmidt Verlag.

Landerl, K. (2017). Learning to Read German. In L. Verhoeven & C. Perfetti (2017). *Learning to Read across Languages and Writing Systems* (pp. 299-322). United Kingdom: Cambridge University Press.

McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E., & Castles, A. (2013). Getting to grips with the heterogeneity of developmental dyslexia. *Cognitive neuropsychology*, *30*(1), 1-24. doi: 10.1080/02643294.2013.784192

McCloskey, M., & Rapp, B. (2017). Developmental dysgraphia: An overview and framework for research. *Cognitive neuropsychology*, *34*(3-4), 65-82. doi: 10.1080/02643294.2017.1369016

Monsell, S. (1987). On the relation between lexical input and output pathways for speech. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production: Relationships between listening, speaking, reading, and writing* (pp. 273–311). London: Academic Press.

Moore, D. M., Porter, M. A., Kohnen, S., & Castles, A. (2012). Detecting different types of reading difficulties: A comparison of tests. *Australasian Journal of Special Education*, *36*(2), 112-133. doi: 10.1017/jse.2012.11

Morfidi, E., Van Der Leij, A., De Jong, P. F., Scheltinga, F., & Bekebrede, J. (2007). Reading in two orthographies: A cross-linguistic study of Dutch average and poor readers who learn English as a second language. *Reading and writing*, *20*(8), 753-784. doi: 10.1007/s11145-006-9035-9

Murphy, K.A., Language and Reading Research Consortium (LARRC) & Farquharson, K. Read Writ (2016) 29: 1745. https://doi.org/10.1007/s11145-016-9651-y

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nelson, J. & Perfetti, C.A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing: An Interdisciplinary Journal*, 29(6), 1069-1103. doi: 10.1007/s111145-015-9616-6.

Palladino, P., Bellagamba, I., Ferrari, M. & Cornoldi, C. (2013). Italian Children with Dyslexia are also poor in reading English words, but accurate in reading English pseudowords. *Dyslexia* 19, 165-177. doi: 10.1002/dys.1456.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. reitsma (Eds.), *Precursors of functional literacy* (pp.189-213). Amsterdam: John Benjamins. doi: 10.1075/swll.11.14per.

Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. doi:10.1080/10888430701530730.

Perfetti, C.A. (2017). Lexical quality revisited. In E. Segers & P. van den Broek. *Developmental Perspectives in Written Language and Literacy* (pp. 51-67). Amsterdam: John Benjamins Publishing Company.

Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, *17*(6), 415-434. doi: 10.1080/10888438.2013.764879

Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 115–158). Berlin: Mouton de Gruyter.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *The Quarterly Journal of Experimental Psychology*, 37, 507–532.

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216. doi: 10.1016/j.cognition.2017.06.024

Sparks, R. L. (2016). Myths About Foreign Language Learning and Learning Disabilities. *Foreign Language Annals*, 49(2), 252-270. doi:10.1111/flan.12196.

Stadie, N., & van de Vijver, R. (2003). A linguistic and neuropsychological approach to remediation in a German case of developmental dysgraphia. *Annals of Dyslexia*, *53*(1), 280-299.

Suárez-Coalla, P., Ramos, S., Álvarez-Cañizo, M. & Cuetos, F. (2014). *Ann. of Dyslexia*, 64 (2), 166-181. doi: 10.1007/s11881-014-0092-5

van der Leij, A., & Morfidi, E. (2006). Core deficits and variable differences in Dutch poor readers learning English. *Journal of Learning Disabilities*, 39(1), 74-90. doi:10.1177/00222194060390010701.

von Hagen, A., Kohnen, S. & Stadie, N. (in preparation). Individual differences in foreign language attainment of children with poor literacy skills.

von Hagen, A., Kohnen, S. & Stadie, N. (under review). Foreign language attainment of children/adolescents with poor literacy skills: A systematic review and meta-analysis.

Wang, H. C., Marinus, E., Nickels, L., & Castles, A. (2014). Tracking orthographic learning in children with different profiles of reading difficulty. *Frontiers in human neuroscience*, *8*, 468. doi: 10.3389/fnhum.2014.00468

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children – Fourth Edition – Australian Standardised Edition*. San Antonio, TX: The Psychological Corporation.

Wight, M.C.S. (2015). Students With Learning Disabilities in the Foreign Language Learning Environment and the Practice of Exemption. *Foreign language Annals*, 48(1), 39-55. doi: 10.1111/flan.12122.

Zoccolotti, P., & Friedmann, N. (2010). From dyslexia to dyslexias, from dysgraphia to dysgraphias, from a cause to causes: a look at current research on developmental dyslexia and dysgraphia. *Cortex*, *46*(10), 1211-1215.

Appendices

Foreign language words used in experimental tasks.

New item nr.	Item	English translation	Grammatical class	Cognate status	Presence of consonant clusters	Length	Time-point of first exposure
1	Hand	hand	Noun	Cognate	Cluster	1 syllable	Early
2	Kreis	circle	Noun	Noncognate	Cluster	1 syllable	Early
3	Bauch	belly	Noun	Cognate	No cluster	1 syllable	Early
4	Bein	leg	Noun	Noncognate	No cluster	1 syllable	Early
5	Drachen	dragon	Noun	Cognate	Cluster	2-3 syllables	Early
6	Dreieck	triangle	Noun	Noncognate	Cluster	2-3 syllables	Early
7	Feuer	fire	Noun	Cognate	No cluster	2-3 syllables	Early
8	Viereck	square	Noun	Noncognate	No cluster	2-3 syllables	Early
9	Wolf	wolf	Noun	Cognate	Cluster	1 syllable	Late
10	Schwein	pig	Noun	Noncognate	Cluster	1 syllable	Late
11	Kuh	cow	Noun	Cognate	No cluster	1 syllable	Late
12	Affe	monkey	Noun	Noncognate	No cluster	1 syllable	Late
13	Elefant	elephant	Noun	Cognate	Cluster	2-3 syllables	Late
14	Schlange	snake	Noun	Noncognate	Cluster	2-3 syllables	Late
15	Ratte	rat	Noun	Cognate	No cluster	2-3 syllables	Late
16	Hexe	witch	Noun	Noncognate	No cluster	2-3 syllables	Late
17	blau	blue	Adjective	Cognate	Cluster	1 syllable	Early
18	gelb	yellow	Adjective	Noncognate	Cluster	1 syllable	Early
19	rot	red	Adjective	Cognate	No cluster	1 syllable	Early
20	schön	beautiful	Adjective	Noncognate	No cluster	1 syllable	Early
21	orange	orange	Adjective	Cognate	Cluster	2-3 syllables	Early
22	prima	great	Adjective	Noncognate	Cluster	2-3 syllables	Early
23	fantastisch	fantastic	Adjective	Cognate	No cluster	2-3 syllables	Early
24	mutig	brave	Adjective	Noncognate	No cluster	2-3 syllables	Early
25	wild	wild	Adjective	Cognate	Cluster	1 syllable	Late
-	klein[a]	small	Adjective	Noncognate	Cluster	1 syllable	Late
26	lang	long	Adjective	Cognate	No cluster	1 syllable	Late
27	dünn	thin	Adjective	Noncognate	No cluster	1 syllable	Late
28	freundlich	friendly	Adjective	Cognate	Cluster	2-3 syllables	Late
29	mittelgross	medium tall	Adjective	Noncognate	Cluster	2-3 syllables	Late
30	lockig	curly	Adjective	Noncognate	No cluster	2-3 syllables	Late
31	gefährlich	dangerous	Adjective	Noncognate	No cluster	2-3 syllables	Late
32	trinken	to drink	Verb	Cognate	Cluster	2-3 syllables	Early
33	springen	to jump	Verb	Noncognate	Cluster	2-3 syllables	Early
34	essen	to eat	Verb	Cognate	No cluster	2-3 syllables	Early
35	heissen	to be called	Verb	Noncognate	No cluster	2-3 syllables	Early
36	schwimmen	to swim	Verb	Cognate	Cluster	2-3 syllables	Early
37	flattern	to flap	Verb	Noncognate	Cluster	2-3 syllables	Early
38	singen	to sing	Verb	Cognate	No cluster	2-3 syllables	Early
39	wedeln	to wag	Verb	Noncognate	No cluster	2-3 syllables	Early
40	schlafen	to sleep	Verb	Cognate	Cluster	2-3 syllables	Late
41	fressen	to eat	Verb	Noncognate	Cluster	2-3 syllables	Late

New item nr.	Item	English translation	Grammatical class	Cognate status	Presence of consonant clusters	Length	Time-point of first exposure
42	beissen	to bite	Verb	Cognate	No cluster	2-3 syllables	Late
43	heiraten	to marry	Verb	Noncognate	No cluster	2-3 syllables	Late
44	klettern	to climb	Verb	Cognate	Cluster	2-3 syllables	Late
45	spielen	to play	Verb	Noncognate	Cluster	2-3 syllables	Late
46	sagen	to say	Verb	Cognate	No cluster	2-3 syllables	Late
47	tanzen	to dance	Verb	Noncognate	No cluster	2-3 syllables	Late

Note. [a]This item was accidentally omitted in the word spelling task.

Supplemental materials

Structured interview with German teacher.

Research project 'Dyslexia and second language learning'

Questionnaire regarding German instruction

Dear German teacher:

In addition to the assessment that we have completed with some of your students in English and German, we would like to ask you some questions about the German instruction you provide. We believe that this complementary information can help us to better understand children´s learning achievement. Thank you very much for your help. As soon as we have the first results of this study, we will share them with you.

Thank you very much!

Language knowledge

1. Is English your native language? **Yes** ☐ No

   If no, which is/are your native language/s? ____________________________________

2. Do you speak another language besides German and English? **Yes** ☐ No

   If yes, which language/s? *A bit of Danish (because of family), Major in French, bits of Japanese and Indonesian*

3. How old were you when you started learning German? *13 years old. At school.*

4. Have you ever lived more than 3 months in a German speaking country? ☐ Yes **No**

   If yes, in which country/ies and for how long? ___________________________

5. How similar is your accent to a native German speaker? :

   ☐ almost native-like X **very** ☐ more or less ☐ not much ☐ not at all

Sound discrimination

1. Do you use the IPA symbols to teach your students the pronunciation of a new word?

   ☐ Yes **No**

2. Do you introduce the spoken and written form a new word simultaneously? ***Yes. From the beginning the written form is shown as well, although the new word is introduced orally. Only from grade 3 on, they learn the alphabet in German and then differences between GPC's are targeted: "thinking with the German or English brain". The idea is to not interfere with GPC´s in English before grade 3. I avoid translations. I always introduce new words paired with pictures, not with translations.***

3. Do you explicitly highlight similarities and differences between German and English sounds (for example that the /r/ is pronounced differently in each language)? **Yes**

4. Do you listen to examples of native speakers' speech with your students (for example Cds, videos, etc.)? ***Yes***.

If yes, with what frequency? ***When it fits in.***

☐ Each class ☐ At least once a week

☐ At least once a month ☐ At least once in each school-term

If yes, what materials do you use? ***Songs, Mutzi series on BBC, native speakers coming in.***

If yes, how long do you listen to such an example in one class?

5 minutes ☐ 10 minutes ☐ 15 minutes ☐ 20 minutes or more

Pronunciation

1. Do you correct your students when they mispronounce a word? ***Yes. If it is a systematic error. On one on one, with teacher model, classmate model.***

2. Do you explain the correct position of the mouth and other articulators to pronounce German sounds that are different form English***?*** ***Yes. Lip reading.***

3. How often do your students speak in German in class on average?
Each class. Start with Guten Morgen when they come into the room.

4. Do you use other strategies to improve your students' pronunciation? ***Yes.***
If yes, which? ***Lipreading***

Vocabulary

1. How many new words do you teach your students on average in a week?
20-30

2. Which of the following strategies do you use, to help your students learn new words?

A) **Children are taught to try to remember a new word by connecting the pronunciation of this word with a known word in their native language. For example the word KUH starts with the same sound /k/, as COW and both are short, one-syllable words. *E.g. Die KUH macht MUH***

B) Children are taught to remember the new word together with the translation into their native language. NO – avoid this!!

C) **Children are taught to remember the new word together with an image of the concept. No translation is involved.** ***Yes, systematically like this! In their books as well.***

D) **Children learn the new word in the context of a text, comic or video. No translation is involved. Children learn the word through a communicative situation.**

E) No specific strategies are taught.

F) This strategy that was not mentioned before is used: ***many language games, concrete objects, kinaesthetic side of learning, work it out by themselves, build on prior knowledge***

3. How often do you test if your students remember the new vocabulary?

more than once a week — **once a week - spelling test**

once very 2 weeks — once a month

4. How do you test if you students remember the new vocabulary?

picture-word matching — translation

fill in the gaps — another form: ___

Reading and spelling

1. Which of the following strategies do you use to hep your students learn to read and write in German (tick one or more boxes)?
   A. Children are taught to remember the word as a whole image. ***No.***
   B. Children are taught connections between letters and sounds, so that they can read and write known words, but also new words. For example tat the letter z is pronounced /ts/. Main strategy
   C. Children are taught rhyme words that are written and pronounced in the same way. For example Haus, Maus, etc. E.G. Australien und Frau – neun und Europa
   D. Children learn that words with a similar meaning are written similarly. For example Baum - Bäume or fahren – Fahrrad. With plural – Apfel, Äpfel
   E. No specific strategy is introduced.
   F. The following strategy that was not mentioned before is used:

2. Do your students read aloud in German class? ***Rare and with patterns – same sentence with varying ending for example***

If yes, how often?

In every class — At least once a week

At least every two weeks — At least once a month

3. How often do your students write in German class?

In every class — At least once a week

At least every two weeks — At least once a month

4. Do you explicitly highlight similarities and differences between the German and English writing system (for example the German <sch> and English <sh>)? ***Yes*** No

5. Do you correct your student´s spelling mistakes? If yes, how? ***Circle the error; sp = check spelling; capital letters are big problem; overgeneralization with Umlaut***

Learning German with dyslexia

1. Have you ever taught German to a child diagnosed with dyslexia? ☐ Yes ☐ No

If yes, which of the following situations have you experienced? ***All. Motivation is important.***

Children with dyslexia have ☐ more/ ☐ the same/ ☐ less difficulties learning German than their peers without dyslexia.

Difficult to say. Often bad communication with special needs teachers and L1 teachers.

2. If you have observed that children with dyslexia struggle learning German, which domains were difficult? ***Very often interference form L1. Not so much specific problem.***

☐ A. Sound discrimination (for example distinguishing two similar words *kam* and *Kamm*).
☐ B. Pronunciation (for example saying /k/ at the end of the word Dach, instead of /χ/).
C. Vocabulary learning.
☐ D. Reading and spelling.
☐ E. General language comprehension (for example understanding instructions).
☐ F. General language production (for example answering to simple questions).

3. Do you think that children with dyslexia should have the possibility to be exempted from learning a second language at school? ☐ Yes ☐ ***No. But it is very common in Australia too.***

4. Do you think that second language instruction should be modified for children with dyslexia?
☐ Yes ☐ No If yes, how?
Possibility of teacher aid helps a lot. Now computer system with learning plan from special needs teacher and each teacher (music, sports, etc.) writes how they can adapt these suggestions to their class.

5. Do you think that assessment of German skills should be modified for children with dyslexia?
☐ Yes ☐ No If yes, how?
____
____

6. Do you think that learning a second language might have negative effects on dyslexic children´s first language literacy skills?
Yes ***No. Many benefits instead.***

If yes, have you ever observed this in a student? In which way? ________________________
____

7. If you had 100.000 Dollars available for a research study concerning dyslexia and second language learning, what study would you propose? ***Nothing specific.***

Coding scheme used to determine cognate status of foreign language words following Kohnert, Windsor & Miller (2004, p.524).

Feature Overlap	Scoring	Example
Initial sound	3 = Same consonant	Drachen – dragon
(0-3 points)	2 = Same vowel	orange - orange
	1 = Similar sound (e.g., same sound class or one of consonant cluster)	schwimmen - swim
	0 = Complete mismatch between initial sounds	klein – small
Number of syllables	2 = Same number of syllables	Kuh - cow
(0-2 points)	1 = Different by only 1 syllable	singen - sing
	0 = Different by more than 1 syllable	schön – beautiful
Consonants	3 = >70% consonant overlap	lang - long
(0-3 points)	2 = 50-70% consonant overlap	Feuer - fire
	1 = <50% consonant overlap	rot - red
	0 = No consonant overlap	mutig – brave
Vowels	2 = ≥80% vowel overlap	Hand - hand
(0-2 points)	1 = 50-80% vowel overlap	Ratte - rat
	0 = no vowel overlap	prima - great

Note. This coding scheme was based on Kohnert, Windsor & Miller (2004, p.524). Each German target word was scored relative to its English translation. Four different features were considered. The maximum score was 10. The items that scored 0-4 points were considered non-cognates, while 5-10 points qualified as cognate

General discussion

The overall aim of this thesis was to investigate individual differences in foreign language attainment of children with poor literacy skills. The common belief that students with poor literacy skills are worse at learning foreign languages than their peers with typical literacy skills was a central motivation to our work. We were concerned that the available evidence supporting this belief was predominantly based on group averages of individual poor readers/spellers with very variable cognitive profiles. As a consequence, we suspected that the findings reported by previous studies were possibly not representative for individual poor readers/spellers and could even be misleading, if taken as a basis to make practical decisions on the foreign language education of individual students. Therefore, we aimed to contribute towards a better understanding of this issue by providing more data on individual differences in foreign language attainment of children with poor literacy skills.

Summary of findings

In **Chapter 1,** we presented a systematic review and meta-analysis looking at how successful children/adolescents with poor literacy skills are in learning a foreign language, as compared to children/adolescents with typical literacy skills. Our results revealed that past research should be interpreted with caution due to at least four of the following reasons. First, information is limited to a small subset of foreign language skills and can therefore not be used to address questions on overall foreign language attainment. Second, the effects reported across different study reports are highly heterogeneous, making it difficult to interpret overall effects. Third, differences related to participant characteristics, foreign language instruction and foreign language assessment between study samples are likely to explain at least some of the observed heterogeneity between study-effects. However, whether heterogeneity was impacted on by these factors could not be tested, as most of the studies did not report relevant information. Fourth, the performance variation within the poor reader/speller group was significantly larger than in the control group for many foreign language measures. This means that we do not know if the group averages, on which this evidence is almost entirely based, are representative of the performance patterns of the individual poor readers/spellers within the group.

In **Chapter 2,** we followed this gap in the literature up by contrasting group and individual results on eight foreign language subskills in German speaking children with poor and typical literacy skills who were learning English as a foreign language at school. In line with past research, group comparisons indicated a lower foreign language attainment for children with poor literacy skills on six out of eight foreign language measures. However, at an individual level, more than half of the poor readers/spellers were just as successful as their peers with typical literacy skills on the eight foreign language measures assessed in this study. Moreover, the poor readers/spellers that did show a lower foreign language attainment, varied with respect to the tasks that they struggled on. This study indicated that there can be a complete mismatch between group and individual comparisons.

Based on the individual variability in the foreign language performance of children with poor literacy skills documented in Chapters 1 and 2, in **Chapter 3** we examined potential sources of this variability. To this end, we investigated cross-linguistic associations between eight native and equivalent foreign language subskills in the same poor readers/spellers that we reported on in Chapter 2. We found native language speech sound perception, spoken word production, nonword reading, word reading and spelling skills to significantly contribute to explaining individual differences in equivalent foreign language measures. Furthermore, only foreign language learning motivation, but none of the other broader cognitive and affective measures, also included in this study, played a significant role in accounting for the observed variance in poor reader/spellers' foreign language performance.

In **Chapter 4** we focused specifically on individual differences in foreign language word knowledge. Analyses focused on individual performance profiles of ten English speaking children with poor literacy skills, who were learning German as a foreign language.

General discussion

Data was collected on six different tasks involving the same 47 German words. Results revealed that foreign language word knowledge was highly variable across individuals and words. The specific tasks in which children struggled the most also differed from child to child. However, for most children we found binding from phonological to orthographic word knowledge (measured by word spelling) to be most difficult. In contrast, binding from phonological to semantic word knowledge (measured by spoken word-picture matching) imposed the least difficulties on most poor readers/spellers. Furthermore, in a similar way as for children with typical literacy skills, poor readers/spellers tended to show higher accuracy on foreign language nouns and adjectives, as opposed to verbs, on cognates, as opposed to non-cognates and on monosyllables, as opposed to bi- and tri-syllables.

Unique contributions of this thesis

Overall, the work presented in this thesis contributes towards a deeper understanding of individual differences in foreign language attainment of children with poor literacy skills in at least four ways. First, we provide the first systematic synthesis of the available evidence addressing the question if children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills. This represents a unique contribution, because the success achieved by poor readers/spellers in different foreign language outcome measures as compared to typical readers/spellers has so far only been investigated in individual studies. We therefore contribute a basis to understand the results of individual studies in a broader context and increase statistical power to estimate overall effects (Borenstein, Hedges, Higgins & Rothstein, 2009). Furthermore, this systematic review provides an overview of the available evidence on this issue which can be useful to researchers (in guiding future studies), as well as parents, teachers/specialists and policy makers (in guiding educational decision making).

Second, we show that while poor readers/spellers, as a group, often perform worse than typical readers/spellers on different foreign language tasks, at an individual level many children with poor literacy skills are just as successful as their classmates with typical literacy skills. This is the first time that the validity of group level results on the foreign language attainment of children with poor literacy skills has been specifically investigated by analysing the individual performance patterns of poor readers/spellers. This work underlines the need to be cautious when interpreting past research that averages the performance of individual children with very variable performances. It therefore sheds new light on the suitability of past research on foreign language attainment in children with poor literacy skills to inform theory as well as practice.

Third, we suggest that when poor readers/spellers experience difficulties in certain foreign language subskills (i.e. speech sound perception, spoken word production, nonword reading, word reading and spelling skills), deficits in equivalent native language subskills can be found due to the reliance on common underlying language resources. This represents a first step to extend our knowledge on the potential sources of individual differences in foreign language attainment of children with poor literacy skills. It extends past research in providing information on the role of native language skills at different stages of foreign language processing in children with poor literacy skills. This information can therefore serve as the basis for examining theories and as a guide to determine which foreign language subskills might be especially difficult for individual child with poor literacy skills.

Lastly, we propose that children with poor literacy skills show variable success in tasks that rely on foreign language word knowledge, because they show different strengths and weaknesses in the underlying cognitive mechanisms involved in foreign language word processing. This is the first time that the underlying cognitive mechanisms involved in foreign language word knowledge have been systematically examined in poor readers/spellers across different tasks with the same set of foreign language words. In this way we deepen our understanding of the possible levels of foreign language word processing that can be especially difficult for poor readers/spellers and thus inform theories and models of foreign language word processing. This again can aid the identification of specific deficits in foreign language word processing and facilitate the implementation of support strategies.

In summary, the work presented in this thesis extends our knowledge on foreign language attainment of children with poor literacy skills and adds towards an evidence base that can guide researchers, parents, teachers and specialists concerned with reaching a better understanding of the foreign language performance of children with poor literacy skills. In the following sections, we provide further details on overall theoretical, methodological and practical implications of our work.

Theoretical implications

The findings presented in this thesis underline the need to draw on theoretical models of language processing to explain the heterogeneity observed in children with poor literacy skills. Being able to read and write are complex cognitive skills comprising multiple underlying components that function in an interconnected manner (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Coltheart, 2006). Deficits in different components or links between components produce different types of poor literacy profiles (Coltheart, 2015; Coltheart & Kohnen, 2012; Friedmann & Coltheart, 2016; McArthur et al., 2013; McCloskey & Rapp, 2017). Models of language processing describe the subcomponents and links that are needed to successfully complete different tasks. They can therefore also be used to identify different levels at which processing can break down and describe different subtypes of language disorders.

In this thesis we relied on Ellis and Young's (1988) model of language processing, because it has (in different version) been extensively applied to capture individual differences in the native language performance of children with poor literacy skills (Friedmann & Coltheart, 2016; Kezilas, Kohnen, McKague & Castles, 2014; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012; Kohnen, Nickels, Geigis, Coltheart, McArthur & Castles, 2018; Sotiropoulos & Hanley, 2017; Stadie & van de Vijver, 2003). By integrating this framework with current models of bilingual processing (i.e. Bilingual Interactive Activation Plus model - Dijkstra & van Heuven, 2002; Bilingual Interactive Model of Lexical Access - Grosjean, 1988, 1997; Revised Hierarchical Model - Kroll & Stewart, 1994), we showed that Ellis and Young's model (1998) can also successfully be applied to describe individual differences in foreign language attainment of children with poor literacy skills. The main advantage of this framework over other models of bilingual processing is that it encompasses input and output levels and is applicable to oral and written language processing at the same time. This allows a more comprehensive analysis of performance patterns across different tasks and hence, gives a more complete overview of individual strengths and weaknesses in native and foreign language processing. While in this thesis we used this framework to investigate individual differences in native and foreign language performance of children with poor literacy skills is likely to also prove useful to describe performance patterns of other bilingual populations (see Gray & Kiran, 2013 for a similar approach to people with bilingual aphasia).

Another implication of this thesis is that we showed that theoretical models of word knowledge need to specify not only the sources of different types of information involved in the representations of a word, but also address how these sources are interconnected. We refer, in particular, to the lexical quality framework by Perfetti and Hart (2002) (see also Perfetti, 2007, 2017). While this framework is very useful to highlight that word knowledge varies from individual to individual and from item to item, it is unclear how different types of word information (i.e. phonological, orthographic and semantic knowledge) are integrated into a coherent representation. Within the lexical quality framework, the links between the so-called 'phonological, orthographic and semantic constituents' of the representation of a word are referred to as 'constituent binding'. However, Perfetti himself expresses that this core feature is relatively under-specified and has not been the focus of past research (p.64, Perfetti, 2017).

In this thesis we propose that language processing models, such as Ellis and Young's (1988) model, can complement this aspect of the lexical quality framework by sharpening the idea of constituent binding. While models of language processing also assume that lexical representations are built via the integration of phonological, orthographic and semantic information, they focus specifically on the processing steps that link different types of information to one another. This seems to be exactly what the lexical quality framework refers to under the term 'constituent binding'. Therefore, future studies investigating constituent binding within the lexical quality framework may wish to draw on models of language processing, such as the one by Ellis and Young (1988).

Methodological implications

A central message of this thesis is that it is imperative to consider individual differences when investigating heterogeneous populations, such as children with poor literacy skills. While in some cases, evidence that is solely based on group averages may fail to capture the performance of some of the individuals in a group, in extreme cases they may not even reflect the performance pattern of a single individual (Nugent, 2006; Smith & Little, 2018). In this thesis we suggested several methodological approaches of how individual differences can be taken into account.

First, in the case of meta-analytic studies, common practice has been to only compute overall effect sizes based on central tendency measures such as the standardized mean difference (*SMD*) between groups. A complementary measure, that has recently proven useful to determine the magnitude of inter-subject variability in meta-analyses in the field of biological evolution and nutrition, is the natural logarithms of the ratio of coefficients of variation (Nakagawa et al., 2015; Senior, Gosby, Lu, Simpson & Raubenheimer, 2016). This measure provides an overall effect of the difference in performance variation across participant groups and can easily be computed based on the same data collected to compute *SMDs* (i.e. *M, SD* and *n*). The natural logarithms of the ratio of coefficients of variation provides important information for the interpretation of results derived from central tendency measures. For instance, in the meta-analysis presented in Chapter 1, only relying on overall *SMDs* would have led to the conclusion that children/adolescents with poor literacy skills show a lower foreign language attainment than their peers with typical literacy skills on many outcome measures. However, the fact that foreign language performance varied more in children/adolescents with poor literacy skills than in control participants emphasizes that this conclusion might not apply to a significant proportion of poor reader/spellers. We therefore recommend that future meta-analyses focusing on heterogeneous populations, such as children/adolescents with poor literacy skills, include computations related to variation of performance in outcome measures to capture individual differences.

Second, in the case of group studies, one way to take individual differences into account is to rely on linear mixed modelling analyses instead of more traditional statistical techniques based on group averages (e.g. Mann-Whitney *U* tests, *ANOVAs*, *t*-tests, regular regression techniques that do not take into account random factors). The advantage of using linear mixed modelling analyses is that the systematic variance resulting from individual differences between participants is taken into account when computing the effect of an independent variable on a dependent variable. In contrast, more traditional techniques are based on averaging over subjects and items and thus, disregard the impact of by-subject and by-item variation (also see Winter, 2013). Another advantage of these types of analyses is that the results can expressed at the group level, but also at an individual level (i.e. showing participant coefficients). This information can then be used to verify if the individual results are consistent with the group-level results or if there are individuals that deviate from group tendencies. We used this approach in the studies reported in Chapters 3 and 4 of this thesis.

A third methodological approach to consider individual differences is to adopt a single case or case series approach and draw on single case statistics, as proposed by Crawford and colleagues (e.g. for a complete overview of these statistical techniques see http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodology.htm). The statistical techniques used in this approach, among other possibilities, allow for an estimation of the extent to which the performance of an individual participant differs from the average performance of a control group (Crawford, Garthwaite & Porter, 2010; Crawford & Howell, 1998). This methodology has widely been used in cognitive neuropsychological approaches to describe deficits in children with developmental cognitive disorders, including children with poor literacy skills (e.g., Friedmann & Lukov, 2008; Kohnen, Nickels, Castles, Friedmann & McArthur, 2012; Träff, Olsson, Östergren & Skagerlund, 2017). By focusing on individuals, rather than groups of participants, as the unit of empirical investigation, case studies are inherently in a better position than group studies to capture individual differences (Castles, Kohnen, Nickels & Brock, 2014; Smith & Little, 2018). However, combined analysis of group averages and individual scores, such as reported in Chapter 2 of this thesis, are also possible and enable more direct comparisons with previous studies that are solely based on group averages (Nickels, Howard and Best, 2012; Nugent, 2006).

Practical implications

This thesis shows that past research in favour of the common belief that children/adolescents with poor literacy skills are worse at learning a foreign language than their peers with typical literacy skills should be interpreted with caution. This is important for educational policies on foreign language instruction for children with poor literacy skills. Also parents, teachers and clinicians should keep in mind that an individual student with poor literacy skills might be just as successful as other students with typical literacy skills. Instead of relying on the false common belief that all poor readers/spellers will struggle in learning a foreign language, foreign language attainment should be closely monitored and support and/or alternative teaching methods put in place when necessary.

A useful way of predicting if an individual student with poor literacy skills might be at greater risk of experiencing foreign language difficulties is to analyse the specific deficits that are visible in the child's native language profile. Information on the child's strengths and weaknesses on certain native language subskills will likely be available from previous assessments and can be used to determine the risk of potential foreign language difficulties. The results of this thesis indicate that poor readers/spellers with deficits in certain native language skills (i.e. speech sound perception, spoken word production, nonword reading, word reading and spelling skills) are likely to experience the same difficulties in equivalent foreign language subskills. This information could therefore be used to support these areas of foreign language learning from the beginning of foreign language instruction.

In some contexts, it might even be possible for specialists to rely on the theoretical framework that we presented in Chapter 3 of this thesis, to analyse cross-linguistic interdependence between a student's native and foreign language performance. Particularly for native German and English speaking children, this could perhaps be done by relying on information from standardized assessments tools that are based on Ellis and Young's (1988) model of language processing and are commonly used to diagnose language disorders (e.g. in German: PhoMo-Kids test battery for children from preschool to third year of primary school - Stadie & Schöppe, 2013; In English: the Castles and Coltheart 2 (CC2) test of single-word reading - Castles, Coltheart, Larsen, Jones, Saunders & McArthur, 2009; the Diagnostic Spelling Test for Irregular Words (DiSTi) - Kohnen, Colenbrander, Krajenbrink & Nickels, 2015; the Diagnostic Spelling Test for Nonwords (DiSTn) - Kohnen et al., 2015).

Another practical implication of this thesis concerns the comparison of poor readers/spellers' performance on the same foreign language words across different tasks. These kind of analyses can provide insight into the underlying cognitive mechanisms that are most difficult for each child. This might be a simple way for teachers to decide, which aspects of foreign language word knowledge the child needs most support in. For example, the findings from this thesis indicate that producing the orthographic form of a foreign language word after having heard the phonological word form (as required in spelling to dictation tasks), might be a specific weakness for many children with poor literacy skills. For these children, it may be beneficial to be exposed to more activities in which the connection from the phonological to the orthographic information of a word could be strengthened. Also, using spelling tests to assess vocabulary knowledge will disadvantage these students. In contrast, accessing the semantic information of a word after having heard the phonological word form, as required in choosing a matching picture after hearing a foreign language word, seems to be a strength for most poor readers/spellers. This information could therefore, for example, be used to support oral text comprehension in communicative situations within the classroom.

Lastly, it might be useful for teachers and specialists to keep in mind the impact of different psycholinguistic variables on foreign language word knowledge, when implementing support strategies for poor readers/spellers with foreign language learning difficulties. While further research is needed in this respect, the evidence presented in Chapter 4 seems to show that the same psycholinguistic variables that have been shown to play a role in learning foreign language words in unselected populations (for an overview see de Groot & van Hell, 2005) are also important in poor readers/spellers. More specifically, verbs seemed to be more difficult than nouns and adjectives, non-cognates seemed to be more difficult than cognates and bi- and tri-syllable words seemed to be more difficult than monosyllabic words. This information can be used to determine different degrees of difficulty in foreign language word sets and for example, guide the selection of foreign language words that should be practiced with more frequency.

Future directions

This thesis makes several unique contributions towards deepening our understanding of individual differences in foreign language attainment of children with poor literacy skills. However, many questions still remain unanswered. In the following sections we outline three possible directions that future research could take.

One direction for future research concerns the investigation of potential sources of individual differences in foreign language attainment of children with poor literacy skills. One of the main contributions of this thesis was to document the variability of foreign language outcomes achieved by poor readers/spellers. While we showed that native language skills play an important role in explaining this variability, more research is needed to deepen our understanding of cross-linguistic associations between native and equivalent foreign language skills of poor readers/spellers. In Chapter 3 of this thesis we reported evidence showing that certain native language subskills of children with poor literacy skills (i.e. speech sound perception, spoken word production, nonword reading, word reading and spelling skills) are associated with equivalent foreign language subskills. However, the possible causal mechanisms underlying these associations remain unknown.

One way of addressing this issue would be to conduct a training study that examines the impact of a native language intervention on the foreign language performance of children with poor literacy skills. For such a study, we would predict foreign language subskills that are thought to rely on common underlying language resources to show a larger improvement as compared to foreign language subskills that are predominantly based on language-specific resources. In a similar way, it would be interesting to assess if foreign language training in subskills that seem to be based on common underlying language resources generalises to equivalent native language subskills in poor readers/spellers (for a similar idea see Abu-Rabia, Shakkour & Siegel, 2014).

A second potential direction for future research focuses on the impact of foreign language instruction on the foreign language attainment of poor readers/spellers. In relation to this, one of the findings that we reported in Chapter 3 was that poor readers/speller with lower native language word reading skills also showed lower foreign language spoken word production skills. Similarly, children with lower native language spoken word production skills achieved a lower performance in reading foreign language words.

General discussion

We speculated that these results could be reflecting the specific vocabulary instruction received by the participants of this study. The foreign language teacher reported that new vocabulary items were introduced in communicative situations. Following this, students were asked to memorize the new English foreign language word together with equivalent German native language translations from written vocabulary lists included at the end of foreign language textbooks. Therefore, this commonly used vocabulary instruction methodology (also see Nation, 2013) relies heavily on written language. Clearly, children with poor native language literacy skills are likely to be at a disadvantage here. We suggested that their poor native language word reading skills caused spoken foreign language vocabulary skills, in the context of this specific teaching methodology. A training study that systematically assesses vocabulary learning outcomes across oral versus written training regimes might be able to confirm whether or not the relationship is causal.

A third potential direction for future research concerns the range of foreign language outcome measures that have so far been used with children with poor literacy skills. While this thesis focused on examining foreign language subskills that are involved in single word processing, it would be interesting for future studies to investigate more complex foreign language subskills, such as for example oral sentence comprehension and production. These subskills play an important role in everyday communication and are therefore also subject of many activities within the foreign language classroom. The results of the systematic review presented in Chapter 1 revealed that there is insufficient evidence available on these measures. More research is therefore needed.

Concluding remarks

In the introduction of this thesis we asked you to imagine that you asked a random group of people to share their general thoughts on providing foreign language instruction to children that have been identified with poor literacy skills in their native language. Now imagine that you asked the same question, but this time, you ask a group of well-informed teachers and researchers. What would they say? Hopefully somebody would say that children with poor literacy skills are a very heterogeneous group and that therefore some children may show difficulties in foreign language learning, but others may be just as successful as their classmates with typical literacy skills. Someone else might add that even if it is likely that some poor readers/spellers will struggle in learning a foreign language, subskills in which they will experience more difficulties are likely to vary from child to child. Another person might request more information on the specific native language deficits experienced by these children, to estimate if they would be at risk of presenting similar deficits in the foreign language they are asked to learn. Overall, everybody would agree that it is impossible to address this question with a 'one size fits all' answer and that the individual situation of each child with poor literacy skills should be analysed in depth before making decisions. This would ensure that the child's future possibilities are as rich as their peers'.

References

Abu-Rabia, S., Shakkour, W., & Siegel, L. (2013). Cognitive retroactive transfer (CRT) of language skills among bilingual Arabic-English readers. *Bilingual Research Journal*, *36*(1), 61-81. doi: 10.1080/15235882.2013.775975

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-analysis*. UK: John Wiley & Sons, Ltd.

Castles, A., Coltheart, M., Larsen, L., Jones, P., Saunders, S., & McArthur, G. (2009). Assessing the basic components of reading: A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties*, *14*(1), 67-88. doi: 10.1080/19404150902783435

Castles, A., Kohnen, S., Nickels, L., & Brock, J. (2014). Developmental disorders: What can be learned from cognitive neuropsychology? *Philosophical Transactions of the Royal Society - B*, 369(1634), 20130407. doi:10.1098/rstb.2013.0407

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, *4*(1), 5-17. doi: 10.1080/13603110600574322

Coltheart, M. (2015). What kinds of things cause children's reading difficulties? *Australian Journal of Learning Difficulties*, 20 (2), 103-112, doi:10.1080/19404158.2015.1114000

Coltheart, M. & Kohnen, S. (2012). Acquired and Developmental Disorders of Reading and Spelling. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp.892-920). New York, NY: Wiley. doi:10.1002/9781118432501.ch43

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108* (1), 204-256. doi: 10.1037/0033-295X.108.1.204

Crawford, J. R. (2018, August 20). *Single-Case Methodology in Neuropsychology*. Retrieved from http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodology.htm

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482-486. doi: 10.1076/clin.12.4.482.7241

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245-260. doi: 10.1080/02643294.2010.513967

de Groot, A. M. B., & van Hell, J. G. (2005). The Learning of Foreign Language Vocabulary. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9-29). New York, NY, US: Oxford University Press.

Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(3), 175-197. doi: 10.1017/s1366728902003012

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.

Ellis, A.W. & Young, A.W. (1988) *Human cognitive neuropsychology*. Hove: Erlbaum.

Friedmann, N., & Coltheart, M. (2016). Types of developmental dislexia. In A. Baron, & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistics perspectives.* Berlin, Boston: De Gruyter Mouton.

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, *44*(9), 1146-1160. doi: 10.1016/j.cortex.2007.09.005

Gray, T., & Kiran, S. (2013). A theoretical account of lexical and semantic naming deficits in bilingual aphasia. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1314-1327. doi: 10.1044/1092-4388(2012/12-0091).

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, 3, 233–274. doi: 10.1080/01690968808402089.

Grosjean, F. (1997). Processing mixed languages: Issues, findings and models. In A.M.B. de Groot, J.F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, (pp. 225-254). NJ: Erlbaum.

Kezilas, Y., Kohnen, S., McKague, M., & Castles, A. (2014). The locus of impairment in English developmental letter position dyslexia. *Frontiers in human neuroscience*, *8*, 356. doi: 10.3389/fnhum.2014.00356

Kohnen, S., Colenbrander, D., Krajenbrink, T., & Nickels, L. (2015). Assessment of lexical and non-lexical spelling in students in Grades 1–7. *Australian Journal of Learning Difficulties*, *20*(1), 15-38. doi: 10.1080/19404158.2015.1023209

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime'becomes 'smile': developmental letter position dyslexia in English. *Neuropsychologia*, *50*(14), 3681-3692. doi: 10.1016/j.neuropsychologia.2012.07.016

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex*. doi: 10.1016/j.cortex.2018.04.008

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

McArthur, G., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., Banales, E., & Castles, A. (2013). Getting to grips with the heterogeneity of developmental dyslexia. *Cognitive neuropsychology*, *30*(1), 1-24. doi: 10.1080/02643294.2013.784192

McCloskey, M., & Rapp, B. (2017). Developmental dysgraphia: An overview and framework for research. *Cognitive neuropsychology*, *34*(3-4), 65-82. doi: 10.1080/02643294.2017.1369016

Nakagawa, S., Poulin, R., Mengersen, K., Reinhold, K., Engqvist, L., Lagisz, M., & Senior, A. M. (2015). Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution*, *6*(2), 143-152. doi: 10.1111/2041-210X.12309

Nation, I.S.P. (2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nickels, L., Howard, D. & Best, W. (2011). On the use of different methodologies in cognitive neuropsychology: Drink deep and from several sources, Cognitive *Neuropsychology*, 28 (7), 475-485, doi: 10.1080/02643294.2012.672406

Nugent, W. (2006). Integrating Single-Case and Group-Comparison Designs for Evaluation Research. *Journal of Applied Behavioural Science*, 32 (2), 209-226. doi: 10.1177/0021886396322007

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. reitsma (Eds.), *Precursors of functional literacy* (pp.189-213). Amsterdam: John Benjamins. doi: 10.1075/swll.11.14per.

Perfetti, C.A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357-383. doi:10.1080/10888430701530730.

Perfetti, C.A. (2017). Lexical quality revisited. In E. Segers & P. van den Broek. *Developmental Perspectives in Written Language and Literacy* (pp. 51-67). Amsterdam: John Benjamins Publishing Company.

Senior, A. M., Gosby, A. K., Lu, J., Simpson, S. J., & Raubenheimer, D. (2016). Meta-analysis of variance: an illustration comparing the effects of two dietary interventions on variability in weight. *Evolution, medicine, and public health*, *2016*(1), 244-255. doi: 10.1093/emph/eow020.

Smith, P. L. & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic bulletin & review*, 1-19. doi: 10.3758/s13423-018-1451-8

Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, *168*, 205-216. doi: 10.1016/j.cognition.2017.06.024

Stadie, N., & Schöppe, D. (2014). PhoMo-Kids. *Phonologie Modellorientiert. Modellorientierte Aufgaben zur Überprüfung phonologischer und dyslektischer Störungen bei Kindern* [Model-oriented phonology. Model-otiented assessment of phonological and dyslexic deficits in children]. Köln: Prolog Verlag.

Stadie, N., & van de Vijver, R. (2003). A linguistic and neuropsychological approach to remediation in a German case of developmental dysgraphia. *Annals of Dyslexia*, *53*(1), 280-299.

Träff, U., Olsson, L., Östergren, R., & Skagerlund, K. (2017). Heterogeneity of developmental dyscalculia: cases with different deficit profiles. *Frontiers in psychology*, *7*, 2000, 1-15. doi: 10.3389/fpsyg.2016.02000

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Appendix

Ethics approval

Universität Potsdam · Am Neuen Palais 10 · 14469 Potsdam

Universität Potsdam
Humanwissenschaftliche Fakultät
Frau von Hagen

Ethikkommission
Vorsitzender
Prof. Dr. Dr. Rapp

Telefon: (03 31) 9 77 17 91
Telefax: (03 31) 9 77 10 89
Datum: 24.02.2017

Endbescheid zum Antrag Nr. 46/2016

Sehr geehrte Frau von Hagen,

die Ethikkommission erhebt keine Einwände gegen das Forschungsprojekt

„Mit Lese-Rechtschreibschwäche Fremdsprachen lernen“ – 1. Amendement zum Antrag 49/2015.

Dem Votum liegt der Beschluss der Ethikkommission 1/53. Sitzung – 20.02.2017 – 9. Telefonkonferenz zu Grunde.

Ich wünsche Ihnen für die Durchführung Ihres Vorhabens viel Erfolg.

Mit freundlichen Grüßen

Prof. Dr. Dr. Rapp
Vorsitzender der Ethikkommission

Bankverbindung:
Landesbank Hessen-Thüringen (Helaba)
Kontonummer: 7110402844
BIZ: 300 500 00
DE09 3005 0000 7110 4028 44
BIC: WELADEDDXXX (Düsseldorf)

Dienstgebäude:
Universitätskomplex I
Am Neuen Palais 10, Haus 9
14469 Potsdam

E-Mail: Nadine.Mohaupt@uni-potsdam.de
Internet: http://www.uni-potsdam.de

Office of the Deputy Vice-Chancellor
(Research)

Research Office
Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 4459
http://www.research.mq.edu.au/
ABN 90 952 801 237

MACQUARIE University
SYDNEY·AUSTRALIA

16 September 2016

Dear Dr Kohnen

Reference No: 5201600544

Title: Dyslexia and Second Language Learning

Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Human Sciences & Humanities)).

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted by:

- Macquarie University

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007 – Updated May 2015) (the *National Statement*).

Standard Conditions of Approval:

1. Continuing compliance with the requirements of the *National Statement,* which is available at the following website:

http://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research

2. This approval is valid for five (5) years, subject to the submission of annual reports. Please submit your reports on the anniversary of the approval for this protocol.

3. All adverse events, including events which might affect the continued ethical and scientific acceptability of the project, must be reported to the HREC within 72 hours.

4. Proposed changes to the protocol and associated documents must be submitted to the Committee for approval before implementation.

It is the responsibility of the Chief investigator to retain a copy of all documentation related to this project and to forward a copy of this approval letter to all personnel listed on the project.

Should you have any queries regarding your project, please contact the Ethics Secretariat on 9850 4194 or by email ethics.secretariat@mq.edu.au

The HREC (Human Sciences and Humanities) Terms of Reference and Standard Operating Procedures are available from the Research Office website at:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics

The HREC (Human Sciences and Humanities) wishes you every success in your research.

Yours sincerely

Dr Karolyn White
Director, Research Ethics & Integrity,
Chair, Human Research Ethics Committee (Human Sciences and Humanities)

This HREC is constituted and operates in accordance with the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (2007) and the *CPMP/ICH Note for Guidance on Good Clinical Practice*.

Details of this approval are as follows:

Approval Date: 16 September 2016

The following documentation has been reviewed and approved by the HREC (Human Sciences & Humanities):

Documents reviewed	Version no.	Date
Macquarie University Ethics Application Form		Received 03/07/2016
Response addressing the issues raised by the HREC		Received 08/08/2016 & 14/09/2016
MQ/CCD School Principal Information Letter	2	10/08/2016
MQ/CCD School Principal Consent Form	2	10/08/2016
MQ/CCD Parent Information Letter (Study 1)	3	14/09/2016
MQ/CCD Parent Information Letter (Study 2)	3	14/09/2016
MQ/CCD Parent Consent Form (Study 1, 2 & 3)	2	10/08/2016
MQ/CCD Teacher Information Letter	2	14/09/2016
MQ/CCD Teacher Information and Consent Form	1	10/08/2016
Tests to be Used (Study 1, 2 & 3)	1	03/07/2016
Motivation Questionnaire (for Students)	1	03/07/2016
Questionnaire Regarding German Instruction (for German Teachers)	1	03/07/2016
Parent Questionnaire	1	03/07/2016
Student Questionnaire	1	03/07/2016

***If the document has no version date listed one will be created for you. Please ensure the footer of these documents are updated to include this version date to ensure ongoing version control.**