JYU DISSERTATIONS 267

Maria Psyridou

Reading Development from Age 2 to 16

Divergent Developmental Pathways and Their Early Identification

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND PSYCHOLOGY

JYU DISSERTATIONS 267

Maria Psyridou

Reading Development from Age 2 to 16

Divergent Developmental Pathways and Their Early Identification

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi Agoran auditoriossa 2
elokuun 28. päivänä 2020 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
in building Agora, auditorium 2, on August 28, 2020, at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2020

Editors
Pekka Mertala
Department of Teacher Education, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8254-6

ISBN 978-951-39-8254-6 (PDF)
URN:ISBN:978-951-39-8254-6
ISSN 2489-9003

ABSTRACT

Psyridou, Maria
Reading development from age 2 to 16: Divergent developmental pathways and their early identification
Jyväskylä: University of Jyväskylä, 2020, 73 p.
(JYU Dissertations
ISSN 2489-9003; 267)
ISBN 978-951-39-8254-6 (PDF)

The overall aim of the current dissertation is to examine reading development over time (from age 2 to 16) and investigate the key factors for the early identification and resolution of reading difficulties. It aims to identify ways to predict the development of reading skills and reading difficulties at different ages with the use of very early skills. In particular, the dissertation seeks to increase understanding of how family risk, cognitive skills, and environmental effects affect reading development and reading difficulties. This dissertation builds on data from the Jyväskylä Longitudinal Study of Dyslexia and The First Steps Study. The results of this dissertation suggested that for most children with reading difficulties, the difficulties are persistent from Grades 1 to 9. However, some participants manifested instability in their reading difficulties, and resolving and late-emerging pathways were identified. The results demonstrate that how reading difficulties are identified affects their stability. In particular, the use of single cut-offs can contribute to false impressions, such as how distinct the reading difficulty groups are. The use of different methodologies such as person- vs. variable-oriented approaches may yield differences in the results. The use of latent profile analysis (a person-oriented approach) revealed higher stability in reading skills compared to the use of cut-offs on continuous variables (a variable-oriented approach). The examination of the longitudinal stability of reading difficulties with the use of cut-offs on continuous variables showed that the use of measures with measurement error and the use of single cut-offs affect the longitudinal stability of reading difficulties identification across two time points by causing misclassifications. Furthermore, the examination of the factors that can be used for the early identification of the differential developmental pathways of reading development suggested multiple risk and protective factors. Family risk for reading difficulties, being a boy, and lower performance in cognitive skills (slow rapid automatized naming, difficulty in reading easy words, and low scores in phonological skills, letter knowledge, number counting, and vocabulary) in kindergarten increased the odds of developing reading difficulties. Factors such as being a girl, and good number counting and vocabulary skills in kindergarten seemed to be good markers for the resolution of reading difficulties. A profile with risk factors did not inevitably lead to reading difficulties.

Keywords: reading fluency, reading comprehension, reading difficulties, development, predictors

TIIVISTELMÄ (FINNISH ABSTRACT)

Psyridou, Maria
Lukutaidon kehittyminen varhaislapsuudesta nuoruusikään: yksilölliset kehityspolut ja kehitystä tukevat tekijät
Jyväskylä: Jyväskylän yliopisto, 2020, 73 s.
(JYU Dissertations
ISSN 2489-9003; 267)
ISBN 978-951-39-8254-6 (PDF)

Väitöskirjan tavoitteena on tarkastella lukutaidon kehittymistä pidemmän ajan kuluessa kuin aikaisemmissa tutkimuksissa (ikävuosina 2–16). Tarkoituksena on tunnistaa tapoja, joilla lasten ja nuorten lukutaidon ja lukemisvaikeuksien kehittymistä voidaan ennakoida tarkastelemalla hyvin varhaisia taitoja. Erityisesti pyritään selvittämään, kuinka perheeseen liittyvät riskitekijät, kognitiiviset taidot ja ympäristötekijät vaikuttavat lukutaidon kehittymiseen ja lukemisvaikeuksiin. Väitöskirja perustuu seuraavissa tutkimuksissa saatuihin tuloksiin: 'Lapsen Kielen Kehitys' ja 'Alkuportaat'. Tutkimuksen tulokset osoittivat, että usein lasten lukemisvaikeudet säilyivät peruskoulun ensimmäiseltä luokalta yhdeksännelle luokalle asti. Joidenkin tutkimukseen osallistuneiden lasten lukemisvaikeuksissa esiintyi kuitenkin vaihtelua: osalla ne korjaantuivat ja osalla ilmaantuivat vasta myöhemmässä vaiheessa. Tulosten mukaan lukemisvaikeuksien tunnistamistapa vaikuttaa arvioihin niiden pysyvyydestä. Erityisesti mittavirhe ja yksittäisten raja-arvojen käyttö voi aiheuttaa vääriä tulkintoja. Mittavirhe, jota kaikkeen testaamiseen liittyy, sekä yksittäisten raja-arvojen käyttö lukemisvaikeuden tunnistamisessa vaikuttavat esimerkiksi siihen, miten pysyväksi lukemisvaikeus arvioidaan kahden ajankohdan välillä., Vaikka mittavirheen ja raja-arvojen vaikutus huomioitiin, tutkimuksessa havaittiin kuitenkin, etteivät lukemisvaikeudet aina ole pysyviä: erityisesti lukutaidon nopeaa kehitystä havaittiin osalla niistä lapsista, joilla oli lukemisvaikeutta ensimmäisillä koululuokilla. Havaittiin myös, että lukutaidon erilaisia kehityspolkuja voidaan tunnistaa jo varhaisessa vaiheessa, ja tutkimus toi esille sekä riskitekijöitä että suojaavia tekijöitä. Pysyviä lukemisvaikeuksia kehittyi todennäköisemmin pojille kuin tytöille sekä niille, joiden perheessä esiintyi lukemisvaikeuksia. Heikommat kognitiiviset taidot esikoulussa (hidas RAN, vaikeus lukea helppoja sanoja, heikot fonologiset taidot, heikko kirjaintuntemus, numeroiden luettelemisen taito ja sanavarasto) lisäsivät myös pysyvien lukemisvaikeuksien todennäköisyyttä. Lukemisvaikeuksien korjaantuminen taas näytti olevan todennäköisempää tytöillä ja niillä, jotka hallitsivat hyvin numerosuoran ja joilla oli laaja sanavarasto. Tutkimuksen perusteella voidaan todeta, että lukemisen vaikeudet eivät ole kaikilla lapsilla pysyviä ja on tärkeää tunnistaa erityisesti niitä tekijöitä, jotka johtavat muutoksiin; joko nopeaan, korjaantuvaan kehityskulkuun tai hitaampaan kehityskulkuun, jossa lukemisvaikeuksia havaitaan vasta myöhemmillä luokilla.

Avainsanat: luetun ymmärtämisen kehitys, lukusujuvuuden kehitys, lukemisvaikeudet, lukutaidon varhaiset ennustajat, lukemisvaikeuden sukuriski

Author

Maria Psyridou
Department of Teacher Education
P.O. Box 35 FIN-40014 University of Jyväskylä
maria.m.psyridou@jyu.fi
https://orcid.org/0000-0003-1425-2279

Supervisors

Docent Associate professor Minna Torppa
Department of Teacher Education
University of Jyväskylä
Jyväskylä, Finland

Professor Anna-Maija Poikkeus
Department of Teacher Education
University of Jyväskylä
Jyväskylä, Finland

Reviewers

Professor Kate Nation
Department of Experimental Psychology
University of Oxford
Oxford, United Kingdom

Professor Oddny Solheim
Norwegian Reading Centre
University of Stavanger
Stavanger, Norway

Opponents

Professor Kate Nation
Department of Experimental Psychology
University of Oxford
Oxford, United Kingdom

ACKNOWLEDGEMENTS

Completing this PhD has been one of the most fascinating adventures of my life. It has allowed me to discover my passion for research, mature my way of thinking and develop my personality in multiple directions. During the last five years, I have had the opportunity to meet and collaborate with amazing people, each of whom I feel the need to thank separately.

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Minna Torppa and Prof. Anna-Maija Poikkeus who gave me the opportunity to complete this PhD under their valuable guidance. None of this work would have been possible without you. Your support, trust, patience, motivation, and insights have been invaluable. I am grateful for all the time you have spent for me and with me. Your positive attitude, and your constructive and positive feedback, inspired me and made me a better researcher. Thank you both!

A special mention goes to Prof. Asko Tolvanen. I wish to thank you for all the extensive instruction about Mplus and all your valuable support with the analyses. Without your support, immense knowledge, and patience, it would not have been possible to carry out this research. Despite the struggles, the analysis was one of the most exciting parts.

I would also like to express my gratitude to all the co-authors of the individual papers: Dr. Kenneth Eklund, Prof. Marjia-Kristiina Lerkkanen, and Prof. Peter de Jong. You have been amazing collaborators. Your feedback improved not only the papers but also me as a researcher. Prof. Peter de Jong gave me the opportunity to visit the Faculty of Social and Behavioural Sciences at the University of Amsterdam. Apart from being a wonderful academic host, he also provided me with amazing support and guidance.

I would also like to thank Prof. Kate Nation and Prof. Oddny Solheim for their time and effort to review this dissertation. Your constructive comments and observations allowed me to improve it. I would also like to thank Prof. Kate Nation for agreeing to act as my opponent in the public examination.

This dissertation utilized data from the Jyväskylä Longitudinal Study of Dyslexia and The First Steps Study. A special thanks goes to everyone who participated in any way. Without their participation, this dissertation could not have been completed.

Last but not least, I feel the need to express my deepest gratitude to my parents, my brother, and to my friends, who have all been by my side all these years, providing me with unfailing support and encouragement, even when I was grouching.

This dissertation was financially supported by the Department of Teacher Education and the Alli Paasikivi Foundation.

Jyväskylä 26.4.2020
Maria Psyridou

LIST OF ORIGINAL PUBLICATIONS

The dissertation is based on the following original publications:

I. Psyridou, M., Eklund, K., Poikkeus, A. M., & Torppa, M. (2018). Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16. *Research in Developmental Disabilities, 78*, 114-124. https://doi.org/10.1016/j.ridd.2018.05.004

II. Psyridou, M., Tolvanen, A., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (2020). Longitudinal stability of reading difficulties: examining the effects of measurement error, cut-offs, and buffer zones in identification. *Frontiers in Psychology, 10*, 2841. https://doi.org/10.3389/fpsyg.2019.02841

III. Psyridou, M., Tolvanen, A., de Jong, P. F., Lerkkanen, M. K., Poikkeus, A. M., & Torppa, M. (under review). Developmental profiles of reading fluency and reading comprehension from grades 1 to 9 and their early identification.

For Studies I, II, and III, the author of the dissertation, taking into account the comments and suggestion of all co-authors, contributed to the statistical analysis, writing of the manuscript, and the review process. The data used in the three publications had been collected as part of two large longitudinal studies, the Jyväskylä Longitudinal Study of Dyslexia and The First Steps study.

FIGURES

TABLE

CONTENTS

1 INTRODUCTION

Many of our activities require reading. For an individual to function properly and independently in modern society, reading is a critically and increasingly important skill. One of the main objectives of education is to teach young children to read and use reading for learning. Children with reading difficulties are at a high risk for various negative consequences concerning their success in education, professional life, everyday activities, and their general participation in society as active citizens. Because the Finnish language is considered highly transparent (Aro, 2017), almost all children can read accurately after one year of formal instruction of reading (Lerkkanen et al., 2004; Soodla et al., 2015). However, even the high transparency of Finnish orthography does not guarantee that all children would acquire reading fluency and reading comprehension consummate to expected age level and, depending on the criteria used, reading difficulties are identified in 5%–20% of children (Lerkkanen et al., 2010). There is growing evidence suggesting that, in contrast to the customary assumption, reading difficulties are not always stable over time for all children. There are some children who develop reading difficulties after Grades 3–4 despite age-appropriate acquisition of early reading skills and others who have age-appropriate reading skills after Grades 3–4 despite reading difficulties during the early grades (Catts et al., 2012; Etmanskie et al., 2016; Leach et al., 2003; Lipka et al., 2006; Torppa et al., 2015). The reasons for poor reading skills as an adolescent as well as those for the resolution of reading difficulties are not yet known. From both theoretical and practical viewpoints, it is vital to examine the development of reading over a long time frame along with the key factors in the early identification of risks and resolution of reading difficulties. Closer examination of these factors can provide a more reliable prediction of children with late-emerging and resolving reading difficulties as well as aid the early prevention of these difficulties. In addition, the examination of the resolving cases may yield significant knowledge on the mechanisms that trigger the protective factors. These protective factors can, in turn, provide a basis for the development of support systems and intervention programs that will benefit children at risk for reading difficulties.

In this dissertation steps aided with novel methodological analytic tools are taken towards the examination of the development of reading skills over time (from age 2 to 16) and the investigation of the key factors for the early identification of the risks and resolution of reading difficulties. The time frame used in this dissertation covers the transition from the early grades in which children learn to read to later grades in which reading is used for learning and to teach other subjects (Chall & Jacobs, 2003). The knowledge gained from longitudinal studies following children's reading skills development over time has practical importance, such as identifying the precursors of reading skills development or identifying signs of delayed or divergent development.

1.1 Reading fluency and reading comprehension development

Reading is a complex skill relying on many linguistic and cognitive capacities (Castles et al., 2018; Nation, 2019; Perfetti & Stafura, 2014; Snowling & Hulme, 2012). Reading comprises decoding, which is typically assessed by the accuracy and speed of reading aloud, and comprehension, which refers to understanding what has been read (Hulme & Snowling, 2013). In order for someone to be considered a skilled reader, that person must be able to decode print and comprehend the meaning of what was read.

The goal of reading is comprehension. According to Nation (2019), comprehension is not a direct replication of the same form and structure of what someone is reading. While listening or reading, people construct a mental model based on their interpretation of the text that exceeds what was expressed specifically. The text is, instead, more of a basis the reader uses to extract information such as the meaning of words, rules of syntax, and build links with their own background knowledge. Based on this information, the reader can make connections and construct meaning (Nation, 2019). Overall, reading comprehension is a multifaceted process during which the mental representation of the text is formed in the reader's memory (Kendeou et al., 2014). Lower- and higher-level cognitive processes are used by the reader in order to construct meaning from the text. Lower-level processes include the translation of written code into meaningful language units. Higher-level processes include the combination of these units into meaningful and coherent mental representation (Kendeou et al., 2014). Both lower- and higher-level processes start to develop before the beginning of formal instruction of reading and they have been found to predict independently reading comprehension skills at a later age (Hogan et al., 2011; Kendeou et al., 2009).

There are many theoretical frameworks that have been suggested for the understanding of reading comprehension development such as the lexical legacy hypothesis (Nation, 2017), the reading systems framework (Perfetti & Stafura, 2014), the direct and mediational inference model (Cromley & Azevedo, 2007) and the construction–integration model (Kintsch, 1988). Although some of these models describe the process of reading comprehension, they do not explain how

it is developed across time. However, the model that has been used most broadly is the simple view of reading (SVR) model (Gough & Tunmer, 1986; Hoover & Gough, 1990). Over the past thirty years, many studies have supported the SVR model (e.g. Catts et al., 2006; Hjetland et al., 2019; Torppa et al., 2016; Tunmer & Chapman, 2012; see García & Cain, 2014 for a meta-analysis). This dissertation is also influenced by the SVR model.

According to the SVR model (Gough & Tunmer, 1986; Hoover & Gough, 1990), reading comprehension (RC) is based on two broad separable components: decoding (D) and linguistic comprehension (LC) (RC = D x LC). Decoding can be identified as the ability to convert print into sound accurately and quickly, and linguistic comprehension as the ability to understand spoken words. Decoding is usually measured as word reading fluency while linguistic comprehension as listening comprehension (Nation, 2019). According to the SVR model, the contributions of decoding and linguistic comprehension change over time, when children shift from alphabetic decoding to reading words fluently, even though the model itself cannot explain this development (Castles et al., 2018; Nation, 2019). In addition, both decoding and linguistic comprehension are necessary to facilitate reading and neither is independently sufficient. Poor decoding skills can lead to difficulties in reading comprehension. In contrast, linguistic comprehension has a greater effect on reading comprehension when decoding skills are strong.

In order for children to learn to decode print, they have to learn the alphabetic principle that symbols represent sounds. Once the children acquire the alphabetic principle they can learn specific grapheme–phoneme associations and apply them to reading (Castles et al., 2018). According to Ehri's phase theory (2005, 2017), children move from the partial alphabetic phase to the full alphabetic phase in which they have better acquired the alphabetic principle, and they can apply it consistently and decode unfamiliar words. Reading accuracy, fluency, and comprehension are closely associated skills, especially in the early stages of reading development, and reading comprehension can develop only after a child has become fluent enough (Florit & Cain, 2011; García & Cain, 2014). Good reading fluency skills can support reading comprehension because well automatized word reading skills reduce the resource demands of cognitive processes (e.g. memory and attention), which can then be devoted to comprehension (Perfetti, 1985).

The relationship between reading fluency and reading comprehension is strong during the early grades and diminishes over time as children become "fluent enough" (Florit & Cain, 2011; Santos et al., 2019; Torppa et al., 2016) although not ceasing to exist (Artelt et al., 2001; Verhoeven & van Leeuwe, 2008). A study conducted among Finnish-speaking children (Torppa et al., 2016) showed that the direct effect of reading fluency on reading comprehension disappears after Grade 1 while the effect of listening comprehension remains for a longer time. Similarly, in a study with English-speaking children in Grades 3, 7, and 10, the effect of decoding (word reading accuracy and fluency) on reading comprehension decreased across grades (from .38 to .06) whereas the effect of

listening comprehension increased (from .59 to .89) (Kershaw & Schatschneider, 2012). In addition, although some studies report predictive links from reading fluency to reading comprehension (Cadime et al., 2017; Kim et al., 2015; Kim et al., 2011; Santos et al., 2019; Tilstra et al., 2009), studies examining bidirectional effects between reading fluency and reading comprehension have shown that they are not equally strong, with the effects from fluency to comprehension to be stronger than the effects from comprehension to fluency (Little et al., 2017).

1.2 Reading difficulties

According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) and the American Psychiatric Association (2013), learning disabilities/difficulties refer to difficulties in reading, writing, or math. Difficulties in reading are further specified into difficulties in accuracy, fluency, and comprehension. For example, such difficulties include inaccurate and slow word reading, and difficulties in understanding the meaning of what is read (American Psychiatric Association, 2013). This specification is in line with the SVR model and research evidence (Catts et al., 2003, 2006; Cirino et al., 2013; Nation, 2019; Torppa et al., 2007), suggesting that not all reading difficulties are alike but different subtypes of reading difficulties can be identified: only decoding difficulties, only reading comprehension difficulties, both decoding and reading comprehension difficulties, and no difficulties. Decoding can additionally be subdivided to reading accuracy and reading fluency or speed.

Difficulties with inaccurate and/or slow reading are referred as dyslexia. In research, dyslexia is usually identified as the lower tail of the continuous, normal distribution of reading accuracy and/or fluency (Branum-Martin et al., 2013; Francis et al., 2005; Shaywitz et al., 1992). In transparent orthographies, such as Finnish, there is a one-to-one correspondence between graphemes and phonemes (Aro, 2017; Lyytinen et al., 2015; Seymour et al., 2003) and learning to read accurately is a fast process. Most children learning to read in transparent orthographies read accurately after one year of formal instruction of reading (Landerl & Wimmer, 2008; Lerkkanen et al., 2004; Soodla et al., 2015). English, on the other hand, is an opaque orthography and learning to read accurately is challenging. In transparent orthographies reading fluency – rather than reading accuracy which is acquired relatively effortlessly or at least fast – is typically considered the main key feature of dyslexia (Frith et al., 1998; Peterson & Pennington, 2012; Wimmer, 1993).

However, difficulties in decoding (accuracy/fluency), do not necessarily mean difficulties also in reading comprehension. In fact, several previous studies have shown a dissociation between decoding and reading comprehension difficulties and indicated that at least average reading comprehension can be achieved even among individuals with inaccurate or slow decoding skills (Catts et al., 2006; Nation et al., 2004; Torppa et al., 2007). In addition, intervention studies have shown that improvement in decoding skills does not necessarily

signify improvement in reading comprehension (Edmonds et al., 2009). In a UK sample, it was reported that 5.3% of the children in primary school and 5% of the same children in secondary school scored 1.5 or more Standard Deviations (SDs) below their peers in reading comprehension despite normal decoding skills (Snowling et al., 2009, as cited in Nation, 2019, p.61; Stothard et al., 2010, as cited in Nation, 2019, p.61). In addition, in a different sample, Nation et al. (2010) found that children with poor reading comprehension skills at the age of 8 had good decoding skills at the beginning of school and they also had age-appropriate reading fluency skill throughout their development. Yet other studies have shown that children with poor decoding skills tend to have poor reading comprehension as well (Ferrer et al., 2015; Nation et al., 2019).

Language difficulties (difficulties in understanding and speaking) may, however, often be co-occurring with decoding difficulties. A recent meta-analysis by Snowling and Melby-Lervåg (2016) indicated that children who later developed dyslexia had language difficulties as infants or toddlers. In a recent study comparing children with dyslexia and/or developmental language disorder at the age of 8 and 9 years, Snowling et al. (2019a) found that although all children with difficulties (dyslexia and/or developmental language disorder) had poor reading comprehension skills at the age of 8, at the age of 9 those with only dyslexia had improved their reading comprehension skills compared to those with developmental language disorder only and those with both difficulties. Similarly, in a Dutch study, vocabulary in Grade 3 explained a large amount of variance in Grade 6 reading comprehension (van Setten et al., 2018) and a Finnish study showed that children with delays in early vocabulary (age 2–5.5 years) had poor reading comprehension skills even up to Grade 9 (Eklund et al., 2018).

Overall, individual differences in reading comprehension development have been associated with several factors, such as cognitive and language skills (e.g. Hjetland et al., 2019; Hulme & Snowling, 2014; Kendeou et al., 2009), educational and family contexts (Dietrich et al., 2006; Puglisi et al., 2017; Senechal, 2006), and motivation and attitudes (Gillett et al., 2011; Petscher, 2010; Yeager & Dweck, 2012). In addition to the cognitive and language skills, molecular genetic studies, twin studies, and family risk studies have shown that family risk seems to have a significant effect on reading comprehension development, and individual differences among children are influenced greatly by genetic effects (e.g. Byrne et al., 2009; Christopher et al., 2015; de Zeeuw et al., 2016; Snowling & Melby-Lervåg, 2016; Taipale et al., 2003).

Many models have been suggested in order to explain how and why children's reading skills differ and have tried to identify factors signifying risk of developing reading difficulties. A useful theoretical framework that attempts to answer these questions is Pennington's (2006) multiple deficit model, which has effectively taken the central place instead of previous models that sought to associate complex developmental disorders (e.g., dyslexia) with a single underlying cause. For a relatively long time, the dominant single deficit model was the phonological deficit hypothesis (e.g. Wagner, 1986; Snowling, 1995).

According to the single deficit models, there is one cognitive deficit that is sufficient and necessary to cause all the behavioral characteristics of a disorder. However, the use of the single phonological deficit model in the case of dyslexia had several weak points. Firstly, there was no single cognitive deficit that could be confirmed across all individuals and could justify all the characteristics of all the cases of dyslexia (Ramus & Ahissar, 2012). In other words, not everyone with dyslexia had a phonological deficit (Pennington et al., 2012; Valdois et al., 2011) and not everyone with a phonological deficit had dyslexia (Snowling, 2008). Secondly, the single deficit model could not explain comorbidity issues. For example, dyslexia has been found to co-occur more often than expected with attention deficit/hyperactivity disorder (Greven et al., 2011; Willcutt et al., 2005; 2007), math difficulties (Landerl & Moll, 2010), and language impairment (Pennington & Bishop, 2009). Consequently, it seems that in the case of complex developmental disorders the single deficit models cannot be supported by research findings.

According to Pennington's multiple deficit model, developmental disorders are typically dependent on the interaction of multiple risk and protective factors whose origin could be environmental or genetic. These risk and protective factors have effects on the cognitive functions that form the basis of typical development, and they contribute to behaviors that define each disorder. Consequently, there is no single factor that is sufficient on its own to cause a disorder but multiple deficits need to be present to produce a developmental difficulty. According to the multiple deficit model (Pennington, 2006), each developmental disorder can be studied at four levels of analysis: etiological, neural, cognitive, and behavioral. Within and between each level there can be interactions. Assuming that the multiple deficit model is accurate, a child who has one or both parents with dyslexia (a) will inherit some of the etiological risk factors, (b) will have higher probabilities of developing dyslexia in contrast with a child whose parents do not have dyslexia, (c) will have deficits in several cognitive skills before learning to read, and (d) even those children who do not develop dyslexia will have lower performance in reading-related cognitive skills compared to those whose parents do not have dyslexia.

An extension of the multiple deficit model is the intergenerational multiple deficit model (van Bergen et al., 2014b) that further specifies the multiple deficit model and also includes the intergenerational transfer of risk and protective factors. In this model, parents' characteristics are also represented in the model. Parental skills can be transmitted to children and affect their reading skills either through genetic pathways or environmentally. Parents provide both genes, as they transfer half of their genes to their children, and the home environment, which can affect children's reading skills. For example, parents with good reading skills might read more, or they might have higher levels of education. Consequently, parental characteristics might be related to children's reading skills. They may provide information about the possibility of a child to develop a reading difficulty or they can act as environmental risk or protective factors.

1.3 Predictors of reading difficulties

The view of reading in the current dissertation is developmental: reading skills build on skills that start to develop years before school entry and, in interaction with the environment, they develop over time. Learning to reading is a continuous and developmental process that starts together and interdependently with oral language development before formal education of reading begins (Lonigan et al., 2000; Whitehurst & Lonigan, 1998). The first step for reading development is the foundation of pre-literacy and language skills during pre-school and kindergarten, which influences later reading skills development (Whitehurst & Lonigan, 1998). It is evident that early pre-literacy and language skills developed in and prior to kindergarten can be strong predictors for later reading skills and reading difficulties. Earlier findings of the key factors for the prediction of reading fluency and reading comprehension development in the domains of early cognitive and language skills, family risk for reading difficulties, parental educational level, and gender are briefly described below.

1.3.1 Cognitive and language factors

In the prediction of the different developmental trajectories of reading fluency and reading comprehension development, specific cognitive and language factors can be informative. Knowledge of the factors underlying reading skills development is essential for the early identification of those in need of extra support as well as for the development of support systems based on individual needs. The specific skills depend on the reading skill in question and the specific orthography being studied. Languages diverge in their orthographic transparency (i.e., in the consistency of the correspondence between graphemes and phonemes). For example, English is an opaque orthography with many inconsistencies, meaning every letter is not always represented by the same sound and, vice versa, every sound is not always represented by the same letter. Finnish, on the other hand, is considered a highly transparent orthography (Aro, 2017), and the consistency between graphemes and phonemes correspondence is close to a hundred percent in both directions, with every letter almost always represented by the same sound and, vice versa, every sound almost always represented by the same letter. Some variations are, thus, likely in the role of specific cognitive skills for decoding print in different orthographies.

Previous studies examining the predictive value of cognitive skills to reading fluency have shown that phonological awareness, letter knowledge, and RAN are good early predictors of reading fluency development across multiple orthographies (e.g., Araújo et al., 2015; Caravolas et al., 2019; Clayton et al., 2019; Landerl & Wimmer, 2008; Melby-Lervåg et al., 2012; Puolakanaho et al., 2008; Snowling et al., 2019). Phonological awareness refers to the conscious ability to identify and manipulate sound units (phonemes and syllables) in oral language (Goswami & Bryant, 1990), and many studies have shown its concurrent and predictive link with reading skills (see Castles & Coltheart, 2004 for a review).

One of dominant views for years was the phonological deficit model and that the underlying cause of poor reading skills of children with dyslexia is a deficit in the phonological processing of sounds (Vellutino et al., 2004). Although there are studies reporting that children with family risk for dyslexia tend to have lower performance on phonological awareness tasks (Boets et al., 2010; Snowling et al., 2003; Torppa et al., 2010; van Bergen et al., 2011, 2012), not everyone with dyslexia has a phonological deficit (Valdois et al., 2011; Pennington et al., 2012) and not everyone with a phonological deficit has dyslexia (Snowling, 2008).

Letter knowledge refers to the ability to link graphemes with the correct phoneme. The predictive role of letter knowledge in reading acquisition and reading difficulties has been established in previous studies (e.g. Puolakanaho et al., 2008, Torppa et al., 2010). The most plausible explanation for letter knowledge being such a strong predictor of reading skills is that the letter–sound knowledge is essential in order for the child to understand the alphabetic principle (i.e., how oral sounds of spoken words are represented by letters in written words; see Byrne & Fielding-Barnsley, 1989). Consequently, the knowledge of grapheme–phoneme correspondence serves as the foundation of learning to read because it can provide children a self-teaching strategy for reading unknown words letter-by-letter (Jorm & Share, 1983; Share, 1995, 1999). In alphabetic languages, decoding (reading accuracy and reading fluency) can be conceived as the link of graphemes and phonemes, and, thus, the role of phonological awareness and letter knowledge is apparent.

Rapid automatized naming (RAN) tasks captures the ability to name quickly visually presented familiar symbols such as objects, letters, colors, or digits (Denckla & Rudel, 1976). Many studies have reported the importance of RAN in reading and its predictive value (Georgiou et al., 2016; Kirby et al., 2010; Landerl et al., 2019; Moll et al., 2014). The reason why RAN and reading fluency are so closely related seems to stem from the shared components that underlie both RAN and fluent reading (Norton & Wolf, 2012). According to Norton and Wolf (2012, p. 429) RAN "invokes a microcosm of the later developing more elaborated reading circuit". However, the specific mechanisms underlying this relation remain unclear. Although phonological awareness, letter knowledge, and RAN have been reported to be strong predictors of later reading fluency skills, their predictive values seem to differ across orthographies. In a recent study (Landerl et al., 2019) comparing phonological awareness and RAN across five different orthographies (English, Dutch, German, Greek, and French), it has been shown that the predictive value of phonological awareness on reading fluency is not the same across the different orthographies and that it increases with the complexity of the orthography. RAN, in contrast, was found to be an equally strong predictor for more and less transparent orthographies (Landerl et al., 2019). Indeed, it has been reported to be a strong predictor of reading fluency across different orthographies (e.g., Finnish: Eklund et al., 2013; Torppa et al., 2016; Norwegian: Lervåg & Hulme, 2009; Estonian: Lukanenok, 2011; Greek: Protopapas et al., 2013; German: Moll et al., 2009; Dutch: de Jong, 2011; English: Compton, 2003).

Early word reading skills is perhaps not surprisingly a good predictor of later reading fluency development. In a Finnish study, initial reading skills were found to predict both reading fluency and reading comprehension in Grades 1 and 2 (Lerkkanen et al., 2004). A recent study (Snowling et al., 2019b) conducted among English-speaking children showed that children with dyslexia and typically developing children differed already at the ages of 4-and-a-half and 5-and-a-half years in their word reading skills, with those who later developed dyslexia having lower word reading skills.

In addition to early word reading skills, phonological awareness, letter knowledge, and RAN, number counting has been reported to be an additional predictor of reading fluency (Koponen et al., 2013, 2016). Number counting tasks capture child's ability to move forwards and backwards in the number line by counting numbers. Two previous studies explored the relation between number counting and reading fluency among Finnish-speaking children and they suggested that number counting is a strong predictor of reading fluency even after controlling phonological awareness, verbal short-term memory (Koponen et al., 2013, 2016), vocabulary, working memory, number concepts, and mother's education (Koponen et al., 2016). Interestingly, in an orthography in which RAN would be considered as the strongest predictor of reading fluency (Georgiou et al., 2008; Landerl et al., 2019; Lepola et al., 2005; Moll et al., 2014), number counting was found to be a more powerful predictor than RAN (Koponen et al., 2016). The reason for the link between reading and number counting could be that both number counting and reading are serial processes, which require monitoring and holding information in one's memory while processing (Koponen et al., 2016). One-by-one processing is essential both in counting and at least in the early stages of reading development (see also de Jong, 2011; Protopapas et al., 2013).

Additional skills are informative for the prediction of reading comprehension. According to the SVR model (Gough & Tunmer, 1986; Hoover & Gough, 1990), reading comprehension relies on decoding and linguistic comprehension. Consequently, children with difficulties in reading comprehension should have difficulties in decoding, in linguistic comprehension, or both. In children with difficulties only in reading comprehension and age appropriate decoding skills, poor decoding is not an explanation. Those with poor reading comprehension skills despite sufficient decoding skills instead have deficits in the other component of the SVR model, linguistic comprehension. Reading comprehension relies on a range of oral language comprehension skills (Hulme & Snowling, 2014), such as expressive and receptive vocabulary (e.g. Florit & Cain, 2011; Nation et al., 2010) and listening comprehension (e.g. Catts et al., 2006; Florit & Cain, 2011; Nation et al., 2010). Nation et al. (2010), for example, reported that those with reading comprehension difficulties at the age of 8 had low oral language skills such as vocabulary and listening comprehension despite age-appropriate phonological skills already at the age of 4-and-a-half years as well across grades in primary school. Similarly, Petscher et al. (2018) found that poor comprehenders in Grade 5 had lower oral language skills at 15,

36, and 54 months. There seems to be a difference in the developmental trajectory of oral language skills between those with only decoding difficulties and those with only reading comprehension difficulties. Although both groups had similar oral language levels at 15 months, those with poor reading comprehension lag increasingly behind their peers from 15 months onward (Petscher et al., 2018). Weaknesses in oral language comprehension skills have been found to manifest before learning to read (Catts et al., 2006; Hulme et al., 2015; Nation et al., 2010; Petscher et al., 2018), thus there could be a possible causal link from early oral language skills on later difficulties in reading comprehension.

1.3.2 Family risk for reading difficulties

Large scale studies have attempted to examine children's reading skills by focusing on environmental factors including parental socioeconomic status, characteristics of the school, and differences between groups of children such as boys and girls (Organisation for Economic Co-operation and Development (OECD), 2010). However, individual differences always prove to be larger than schooling effects and vast differences in reading skills can be observed even between children attending the same school and having the same teacher. Although some of the differences can be explained by environmental factors, it is well established by studies that reading difficulties run in families and differences among children in reading skills are mainly due to genetic factors (e.g., Christopher et al., 2015; de Zeeuw et al., 2016; Olson & Byrne, 2005; Swagerman et al., 2017).

The effects of genetic transmission of reading ability has been confirmed, for example, by recent twin studies both in transparent orthographies such as Dutch (e.g., de Zeeuw et al., 2016; Swagerman et al., 2017) and less transparent ones such as English (e.g., Christopher et al., 2015). Given the foundational role of pre-literacy skills developed during pre-school years and kindergarten, a number of behavioral genetic studies have tried to increase insight into the genetic factors that influence the development of pre-literacy skills and their link with reading development later on (e.g., Christopher et al., 2015; Hart et al., 2009; Samuelsson et al., 2005; Soden-Hensler et al., 2012). Hart et al. (2009), in their twin study, focused on expressive vocabulary and reported significant effects from the genetic as well as the shared environment. In addition, studies examining the effects of genetic and environmental factors on reading fluency and reading comprehension have reported that reading fluency, during the early grades, is strongly affected by genetic factors and the influence of the environmental factors was relatively weak (Byrne et al., 2007; Christopher et al., 2015; Hart et al., 2010; Taylor & Schatschneider, 2010). Reading comprehension, on the other hand, was influenced by genetic factors and the environmental effects were low and/or non-significant (Keenan et al., 2006; Logan et al., 2013).

In line with previous twin studies, studies focusing on family risk for reading difficulties have shown that when one or both of the parents have reading difficulties there is heightened probability that the child will develop reading difficulties. Overall, studies have shown that there is a strong link

between children's reading skills and parental reading skills (Puolakanaho et al., 2008; Torppa et al., 2010, 2011, 2015; Snowling et al., 2003; van Bergen et al., 2012, 2014a, 2015, 2017). In a Finnish study that included children with family risk for reading difficulties, Torppa et al. (2011) found that parental reading skills could predict children's reading skills in Grade 3 even after controlling for children's pre-literacy skills. A meta-analysis (Snowling & Melby-Lervåg, 2016) reported that 29%–66% of those having family risk will develop reading difficulties. In addition, in a Dutch study, van Bergen et al. (2017) reported that fathers' and mothers' reading fluency explained, independently, similarly large amounts of the variance of children's reading fluency. Furthermore, in a Norwegian study, family risk for reading difficulties has been found to predict children's later reading difficulties over and above cognitive factors, parental education, and home literacy environment (Esmaeeli et al., 2019).

Previous studies have suggested that children with family risk for reading difficulties have early difficulties in phonological awareness, RAN, letter knowledge, and vocabulary (Boets et al., 2010; Eklund et al., 2018; Snowling et al., 2003, 2007; Snowling & Melby-Lervåg, 2016; Torppa et al., 2010; van Bergen et al., 2011, 2012). Snowling et al. (2003, 2007) found that even children with family risk for reading difficulties but who do not develop reading difficulties show similar deficits than those children with family risk and reading difficulties. Children with family risk and no reading difficulties performed significantly poorer in letter knowledge and rhyme tasks in kindergarten in comparison to children with no family risk and no reading difficulties (Snowling et al., 2003). In a follow-up study, Snowling et al. (2007) found that at the age of 8 years, 66% of the children with family risk for reading difficulties had developed reading difficulties. Those with family risk and no reading difficulties seemed to compensate their poor decoding skills with good language skills but yet they had significantly poorer reading fluency skills at the age of 13 than did those with no family risk and, on timed tasks, they were as slow as those with family risk and reading difficulties. Similarly, in a Dutch sample, children with family risk and no reading difficulties scored higher in reading accuracy and reading fluency in Grade 2 than did those with family risk and reading difficulties but their performance was poorer than that of the children with no family risk and no reading difficulties (van Bergen et al., 2012). In a different Dutch sample, however, those with family risk and no reading difficulties performed poorer than did those with no family risk and no reading difficulties in non-word reading fluency in Grades 1, 2, and 5, but there were no differences between the groups in word reading fluency in Grade 5 (van Bergen et al., 2011). Yet a Finnish study (Torppa et al., 2010) found no significant differences between those with family risk and no reading difficulties and those with no family risk in pre-literacy skills. No differences among these groups were also observed in Grade 2 in reading accuracy or fluency (Torppa et al., 2010).

As it seems from the above-mentioned studies, most have found a continuum in genetic liability of dyslexia. In other words, those with family risk and without dyslexia perform better than those with family risk and dyslexia but yet poorer than those with no family risk and no dyslexia, meaning that those

with family risk and without dyslexia have higher liability than those with no family risk and no dyslexia, which would fit the hypotheses of the multiple deficit model (Pennington, 2006). However, many factors among the studies differ in order to draw conclusive results. Apart from the different orthographies, the ages and the assessment tasks, one of the reasons for these differences could be the definition of reading difficulties. In studies by Snowling et al. (2003, 2007), reading difficulties, and more specifically dyslexia, is defined as having reading accuracy difficulties and consequently their difficulties are more related with a deficit in phonological awareness and they may miss those with a deficit in RAN. In the Dutch (van Bergen et al., 2011; 2012) study, dyslexia is defined as having difficulties in reading fluency and a deficit in RAN is more likely and they may miss those with a deficit in phonological awareness. However, in the Finnish study (Torppa et al., 2010), both reading accuracy and reading fluency are assessed and consequently all children with deficits either in phonological awareness or in RAN are included.

1.3.3 Gender

Several studies have examined the effect of gender on reading achievement but the results are mixed. On the one hand, there are studies showing that gender does not affect reading achievement; they have identified no significant differences among boys and girls in reading achievement (White, 2007), in the frequency of reading difficulties (Jimenez et al., 2011; Moll et al., 2014) or the identified differences are very small (see Lietz, 2006, for a meta-analysis). On the other hand, there are studies showing that girls outperform boys (Berninger et al., 2008; Clinton et al., 2014; Hawke et al., 2009; Quinn & Wagner, 2015; Rajchert et al., 2014; Rutter et al., 2004; Stoet & Geary, 2013) and that more boys than girls identified with reading difficulties (Hawke et al., 2007; Quinn & Wagner, 2015; Stoet & Geary, 2013). A recent study using Finnish-speaking adolescents in Grade 9 (age 15–16) showed that the risk of scoring in the lowest 10% in reading fluency was 4.4 times higher for boys than girls while for reading comprehension the risk was 2.4 times higher for boys than girls (Torppa et al., 2018).

Differences among boys and girls have also been reported in studies examining the stability of reading difficulties over time, with girls being over-represented in the resolving group and under-represented in the late-emerging group. In the resolving group, 80% of the children were girls but in the late-emerging group only 22% of the children were girls (Torppa et al., 2015). Although the prevalence of boys and girls in the group with stable reading difficulties (reading difficulties in both Grade 2 and Grade 8) was similar (45% of the children were girls), under-representation of girls in the late-emerging groups and their over-representation in the group with resolving reading difficulties shows that in Grade 8 there were more boys than girls with reading difficulties. Similarly, in an English-speaking sample (Etmanskie et al., 2016), it was found that the prevalence of boys and girls was uneven, with 28% of the children with late-emerging reading comprehension difficulties being girls and 72% being boys. Gender differences have also been reported in the Programme

for International Student Assessment (PISA), with girls performing better in reading tasks than boys did in all 32 countries (OECD, 2011, 2013, see also Chiu & McBride-Chang, 2006). Nonetheless, the extent of gender differences varied considerably. Although numerous studies have reported differences in reading achievement and the frequency of reading difficulties among boys and girls, genetic studies have not identified the reasons for these gender differences (Hawke et al., 2006, 2007).

Overall, there are two views that are associated with the origin of the gender differences in reading difficulties. The one view refers to the ascertainment bias; that is, more boys than girls are referred for evaluation of their reading skills. This could occur if more boys than girls express disruptive behavior related to problems with reading. Generally, both teachers in classroom settings and parents are likely to detect disruptive behavior and then the children are referred for evaluation (Prior et al., 1995; Shaywitz et al., 1990). The other view is that indeed reading difficulties are more frequent in boys than they are in girls and boys are more vulnerable (Liederman et al., 2005; Rutter et al., 2004).

1.3.4 Parental education

Numerous studies have explored the effects of parental educational level on children's reading skills. Indeed, many of them have found significant correlations between parental education and children's reading skills (Kikas et al., 2018; Kiuru et al., 2013; Manolitsis et al., 2011; Segers et al., 2016; Torppa et al., 2006; van Bergen et al., 2017; van Setten et al., 2018). A possible explanation for this association could be that parent's education level influences the way they communicate with their children during learning activities (Eccles, 2005). It can also influence the quality of reading activities that parents and children do at home. Previous studies have reported that there is an association between low parental education and a less rich literacy and numeracy environment (e.g., shared reading experiences, literacy resources, parental teaching of reading) (Guo & Harris, 2000). Moreover, higher level of parental education has been related with a richer language home environment, which can result in better children's vocabulary (Mol et al., 2008). On the other hand, less rich literacy environments have been linked with lower language and vocabulary skills (Scarborough & Dobrich, 1994; Torppa et al., 2006). In addition, in parents with higher education levels, their interactions with their children during storybook reading have been associated with better phonological awareness skills in children (Korat et al., 2007; Korat, 2009). These kinds of interactions have been suggested to act as predictors of reading development during the early grades (Sénéchal & LeFevre, 2002).

Given though that reading difficulties are highly heritable, these associations between parental education level and children's reading skills could be masked genetic effects that affect both parents' and their children's reading skills and via reading difficulties their educational pathways and the home environment (Puglisi et al., 2017; van Bergen et al., 2017). In other words,

previous studies have found that the link between the home literacy environment and children's reading skills might be explained (even a part of it) through genes parents share with their children. Parents share their genes with their children but they also shape the home environment. Therefore, a parent may have the genetic predisposition to reading difficulties, which may transfer to the child, and may also be less inclined to buy books and offer a literacy-rich environment. In support of this notion, van Bergen et al. (2017) reported that when parental reading skills are taken into account the significant correlations between children's reading skills and parental education become non-significant. Similarly, van Setten et al. (2018) showed that parental education level was associated with parental reading skills and the effect of parental education level on children's reading fluency skills in Grade 6 disappeared when family risk for reading difficulties was taken into account.

1.4 Stability of reading difficulties across time

Previous studies focusing on reading development are numerous (e.g., Snowling, 2008; van Bergen, 2011), but they have considerable limitations. First, they have mainly focused on reading development until the early grades and follow-up studies beyond the early grades are scarce. The data used in this dissertation enable the analysis of reading development up to Grade 9, which provides the opportunity to address questions related to reasons behind adolescent poor reading skills. Second, most previous studies have focused on the group of children with early identified reading difficulties and assumed persistence in reading difficulties. Nonetheless, the few studies that have investigated the development and stability of reading difficulties beyond the early grades have shown that reading difficulties are not always persistent over time (Catts et al., 2012; Etmanskie et al., 2016; Leach et al., 2003; Lipka et al., 2006; Torppa et al., 2015). Though many children with reading difficulties show persistent reading difficulties across time, many also manifest reading difficulties after Grades 3 and 4 despite normal early reading skill development (late-emerging reading difficulties). There are also some studies that have reported that there is a group of children whose reading difficulties resolve over time (resolving reading difficulties) (Catts et al., 2012; Torppa et al., 2015). These unstable groups are of considerable interest because they can increase our understanding on the developmental risk and supportive mechanisms in reading difficulties. Such information is beneficial for more accurate identification of children with reading difficulties and for designing intervention programs for those having reading difficulties.

However, the methodology used in these studies (Etmanskie et al., 2016; Leach et al., 2003; Lipka et al., 2006; Torppa et al., 2015) causes severe uncertainty whether the unstable groups indeed exist, or they are just the effect of measurement error. For the examination of the stability of reading difficulties across time, these studies have used cut-off points on raw variables with

measurement error included. In general, a practical tool for the identification of reading difficulties is setting a cut-off point in the reading ability distribution. There are two main problems with this practice though. Firstly, the use of cut-off points would be justified if we did not have a normal skill distribution, but this is not the case for reading (Francis et al., 2005). Secondly, it leads to uncertainty regarding the results of the above mentioned studies because of the effect of measurement error (Branum-Martin et al., 2013; Francis et al., 2005; Schatschneider et al., 2016). In particular, as measurement error is inevitable in all reading tests setting a cut-off point based on raw variables necessarily causes misclassifications. If a child's score is just above the cut-off, the measurement error could drop the observed score below the cut-off point, and that child will be falsely classified as having reading difficulties. Consequently, measurement error could have an effect on the accurate identification of those with reading difficulties. Moreover, when the stability of reading difficulties is examined with this methodology the existence of the unstable groups is questionable because changes in someone's reading difficulty status can occur because of true change or because of the influence of measurement error at either or both time points.

It is also likely that the use of single cut-offs could contribute to faulty or biased estimations on the prevalence of the instability of reading difficulties if those belonging to the unstable groups are scoring just above or just below the cut-off point. In addition, it can lead to faulty impressions about how distinct the reading difficulty groups are as well as about the risk children have to develop reading difficulties. Another problem with the use of a single cut-off is its arbitrariness. There is no common consensus on the cut-off point that should be used. Different research groups use different cut-off points, such as the lowest 10% of the reading skill distribution, the lowest 25%, 1 SD below the mean, and 1.5 SDs below. But what should be used in practice, in real life? In addition, even if measurement error was not causing any misclassifications, where and how we set a single cut-off has ethical implications for schooling and placement decisions. It can determine, for example, who participates in an intervention, or who receives extra support and who does not. What happens with those scoring just above the cut-off point? Shankweiler et al. (1999) suggested that a buffer zone could better divide those with and without reading difficulties compared to the use of a single cut-off point. Consequently, buffer zones could be used around the single cut-offs as a precaution against biased conclusions (Shankweiler et al., 1999).

1.5 Aims of the research

The overall aim of this dissertation is to examine reading development over time (from age 2 to 16) and investigate the key factors for the early identification and resolution of reading difficulties. It aims to identify ways for predicting the development of reading skills and reading difficulties at different ages with the use of very early skills. In particular, it seeks to increase understanding of how

family risk, cognitive skills, and environmental effects affect reading development and reading difficulties over a long period. The knowledge gained from longitudinal studies following children's reading skills development over time has practical importance such as identifying precursors of reading skills development or identifying signs of delayed or divergent development. In addition, by gaining knowledge about how reading skills develop across time and which factors affect them more, we can make hypotheses about how to enrich the environments in which learning processes take place as well as to build more efficient support systems based on children's individual needs. The data used in the current dissertation offer the unique possibility to examine reading in a long time frame which covers the transition from the early phases of teaching reading to later phases in which reading is used as a tool for learning and teaching other subjects and texts are increasingly more complicated. Therefore, it offers the possibility to examine long-term development and the effects of early risk and protective factors.

Study I focused on very early expressive and receptive vocabulary, which was assessed at the age of 2 and 2-and-a-half years. The aim was to explore if family risk and expressive and receptive vocabulary within that age range predict reading fluency and reading comprehension development in Grades 2 to 9. Possible differences between the identified groups in their later reading fluency and reading comprehension skills would suggest that delays in early expressive and/or receptive vocabulary could be a risk factor and an early mark for later reading difficulties, even until adolescence.

In Study II, the stability of reading difficulties across primary school (Grades 2 to 6) was the focus of interest. In the center of attention were the problems and the limitations that earlier studies have had in the identification of reading difficulties across time by using small samples, variables with measurement error, and arbitrary cut-offs. Previous studies had shown that reading difficulties are not persistent over time for all children, but there are two other subtypes: late-emerging and resolving groups. However, the methodological problems these studies had, cause severe uncertainty if these unstable groups truly exist. This study examined whether the unstable groups can be identified with the use of more advanced methodology (with the use of simulations) and a larger sample. In addition, it examined how different cut-offs affect the identification of the children with reading difficulties.

Study III examined the developmental profiles of reading fluency and reading comprehension from Grades 1 to 9. It also looked at the predictive value of several cognitive skills, gender, mother's and father's education level, and parental reading difficulties causing increased risk for reading difficulties for the child. Possible differences among the profiles in the examined factors could provide important information on the specific characteristics of each developmental profile as well as factors that could facilitate the early identification of the different developmental trajectories of reading skills.

2 METHOD

2.1 Participants

This dissertation builds on data from two unique large Finnish longitudinal studies: 1) the Jyväskylä Longitudinal Study of Dyslexia (JLD) and 2) The First Steps Study. These studies provide unique possibilities because of the several assessments across time and the repeated measures they include. The first article utilized data from the JLD study, which offered the possibility to examine very early identification of language problems (from 2 years onwards). The other two studies utilized data from The First Steps Study, which allowed longitudinal modeling of reading development and reading difficulties across Grades 1–9 in a large sample.

The JLD is a prospective longitudinal family risk study following children with and without family risk for dyslexia. In the study, 200 children participate and half of them have family risk for reading difficulties, meaning that at least one of the parents has a diagnosis for reading difficulties while the other half comes from families without such a risk. The participants were selected from among families of 9,368 newborns born in the province of Central Finland between April 1993 and July 1996. The selection of the parents with dyslexia followed a three-step procedure. The first step included a short parental questionnaire with three questions regarding difficulties in learning to read and spell among themselves and their close relatives. The second step included a more comprehensive questionnaire, focusing on demographic information and the manifestation of the parents' reading and writing difficulties during childhood, adolescence and among relatives. Those parents who fulfilled the criteria of inclusion in the first and second step were invited for an interview and for an assessment of their reading and writing skills to confirm their present status of dyslexia (for full details of recruitment, see Leinonen et al., 2001).

The criteria for a child to be included in the family risk group (FR, n = 108) included that one of the parents was identified as having reading difficulties in the reading and spelling tests, literacy problems during early school years, and at least one first-degree relative with corresponding difficulties. In the control group without family risk, neither parent had reported reading difficulties or a

family history of dyslexia, and neither had difficulties in reading and spelling tasks (NR, n = 92). The educational backgrounds of the parents were assessed using a 7-point scale: 1 = comprehensive school without any vocational education, 2 = comprehensive school with short-term vocational courses, 3 = comprehensive school with a vocational school degree, 4 = comprehensive school with a vocational college degree, 5 = comprehensive school with a lower university degree (Bachelor's) or a degree at a polytechnic, 6 = upper secondary school with a BA degree or a degree at a polytechnic, and 7 = upper secondary school with a higher university degree (master's or a doctorate degree). The mean of mothers' education was 4.18 (SD = 1.48) in the FR group and 4.52 in the control group (SD = 1.35). The mean of fathers' education was 3.68 (SD = 1.27) and 3.80 (SD = 1.40), respectively. There were no differences in parental education or in their nonverbal IQs between parents in the FR and the control groups. The IQ of all parents, assessed with the Raven B, C and D matrices, was higher than 80 (Raven, Court, & Raven, 1992).

The First Steps Study is a Finnish longitudinal study, following approximately 2,000 children from kindergarten to Grade 9. The follow-up started in 2006 when children (n = 1,880) attended kindergarten (in Finland the fall of the year they turn six years). The follow-up is currently funded to continue until 2020 as the School Path study (up to the end of secondary education). The project investigates the development of academic skills, motivation and engagement, social skills, peer relations, and well-being, parenting, and teacher-student interaction. The study comprises a population-based sample with the 1,880 children followed from kindergarten onwards and their classmates (approximately 2000 children at each timepoint) assessed in Grades 1, 2, 3, 4, 6, 7, and 9. Children's cognitive skills were assessed in kindergarten during fall and/or spring and their reading fluency and reading comprehension skills were assessed in Grades 1, 2, 3, 4, 6, 7, and 9. The sample was drawn from four municipalities: two in central, one in western, and one in eastern Finland. In three of the municipalities, the participants form the whole age cohort of children, and in the fourth the participating children comprise about half of the age cohort. One municipality was mainly urban, one mainly rural, and two included both urban and semi-rural environments. Of the parents who were contacted, 78%–89%, depending on town or municipality, agreed to participate in the study. Of the children's mothers, 7.6% had no education beyond secondary school, 30.2% had a vocational school degree, 23.8% a vocational college degree, 9.9% a bachelor's degree, 24% a master's degree, and 4.6% a doctoral degree. Of the children's fathers, 7.9% had no education beyond secondary school, 33.2% had a vocational school degree, 23.7% a vocational college degree, 9.9% a bachelor's degree, 19% a master's degree and 6.3% a doctoral degree. The distribution of parental education was very close to the national distribution of Finland (Statistics Finland, 2007).

2.2 Research ethics

The data collection and the ethical protocol of the two longitudinal projects, the Jyväskylä Longitudinal Study of Dyslexia and The First Steps Study were both approved by the Ethics Committee of the University of Jyväskylä. In addition, the Jyväskylä Longitudinal Study was also approved by the hospital district of Central Finland as the children participated as newborns in the electroencephalography (EEG) assessments at the maternity ward.

Ethical principles and responsible code of conduct have been followed throughout all phases of both studies. The participation in both studies was voluntary, and all participants had been informed about their right to interrupt their participation in the studies at any time. At the beginning of the studies, the parents of the families provided informed written consents for their and their child's participation. All the assessments were safe for the participants. Extra attention was given to managing risks especially in the childhood years.

The data were stored securely. All activities were designed to guarantee the participants' anonymity and following the principles of data protection. All data have been pseudonymized to maintain privacy and the identity of individual subjects was not known to researchers managing and analyzing the data.

I had access only to the data that were needed in the three studies, and I have signed a contract for using the data. The handling of the data was very cautious and no one else had access to these data. In my research utmost care, integrity, meticulousness, and accuracy has been observed in all phases of the dissertetion from data management to analysis, and the reporting of the results, and ethical aspects have also been reported. In addition, none of the participants can be identified from my results and my papers.

2.3 Measures

Table 1 shows the measures used in Studies I–III and the assessment ages. In all three studies the design was longitudinal. More details about the measures can be found in the original papers.

Study I utilized data from the JLD study. Data were used from children's assessments at the ages of 24 months, 30 months, 8–9 years (Grade 2), 9–10 years (Grade 3), 14–15 years (Grade 8) and 15–16 years (Grade 9). Children's language skills at ages 24 and 30 months were assessed individually by trained testers in a laboratory setting. Reading fluency and reading comprehension in Grades 2, 3, 8, and 9 were assessed via group-administered tests in the classrooms. Reading fluency was assessed with three tasks: oral text reading, oral pseudoword text reading and oral word list reading. Reading comprehension was assessed using a short passage reading comprehension task (Grades 2 and 3) and PISA reading comprehension (Grade 9).

Studies II and III utilized data from The First Steps Study. The assessments conducted in kindergarten (fall 2006 and/or spring 2007), Grade 1 (spring 2008), Grade 2 (spring 2009), Grade 3 (spring 2010), Grade 4 (spring 2011), Grade 6 (spring 2013), Grade 7 (spring 2014) and Grade 9 (spring 2016).

Table 1 Measures and children's ages at assessments in Studies I–III

Measures	Child's age/grade	Study I[1]	Study II[2]	Study III[2]
Language and cognitive skills				
Vocabulary	2 y	x		
	2.5 y	x		
	6 y			x
Phonological awareness	6 y			x
Letter knowledge	6 y			x
Rapid naming	6 y			x
Number counting	6 y			x
Word reading	6 y			x
Listening comprehension	6 y			x
Parental factors				
Parental education				x
Family risk		x		x
Reading skills				
Reading fluency	Grade 1			x
	Grade 2	x	x	x
	Grade 3	x		x
	Grade 4			x
	Grade 5			
	Grade 6		x	x
	Grade 7			x
	Grade 8	x		
	Grade 9			x
Reading comprehension	Grade 1			x
	Grade 2	x	x	x
	Grade 3	x		x
	Grade 4			x
	Grade 5			
	Grade 6		x	x
	Grade 7			x
	Grade 8			
	Grade 9	x		x

[1] Data from JLD follow-up; [2]Data from The First Steps follow-up

3 OVERVIEW OF THE ORIGINAL STUDIES

3.1 Study I

Reading Outcomes of Children with Delayed Early Vocabulary: A Follow-up from Age 2 to 16

The aim of the first study was to conduct a long-term follow-up to examine whether early expressive vocabulary delay (late-talking) can predict reading fluency and reading comprehension development in Grades 2 to 9. We examined further if the prediction differs with the presence of family risk for dyslexia and early receptive vocabulary delay. The study extended the already existing literature by examining the development of reading skills of children with early expressive vocabulary delay longer than previously (up to 16 years), by using a Finnish sample, and by examining the development in relation both to reading fluency and reading comprehension. The existing research on the development of reading skills of children with delays in early vocabulary is nearly entirely limited to findings from studies conducted in the English language context and during the early grades of primary school.

The research questions of this study were as follows: (1) Does early expressive vocabulary delay predict difficulties in reading fluency or reading comprehension in Grades 2, 3, 8, and 9? (2) Is the relationship between early expressive vocabulary delay and reading development different in the presence and absence of receptive vocabulary delay and family risk for dyslexia?

This study utilized data from the JLD study, which offered the possibility for very early identification of language problems/deficits and thorough assessment of family risk for dyslexia. The sample consisted of 200 Finnish-speaking children, of which 108 had family risk for dyslexia and 92 came from families without reading difficulties. The reading development of five subgroups was compared: (1) family risk for dyslexia and no vocabulary delay, (2) family risk for dyslexia and early expressive vocabulary delay, (3) family risk for dyslexia and early expressive and co-existing receptive vocabulary delay, (4) no

family risk for dyslexia and early expressive vocabulary delay, and (5) no family risk for dyslexia and no vocabulary delay. Expressive and receptive vocabulary skills were assessed at the age of 2–2.5 years, and reading skills in Grades 2, 3, 8, and 9 (ages 8 to 16) (see Figure 1).

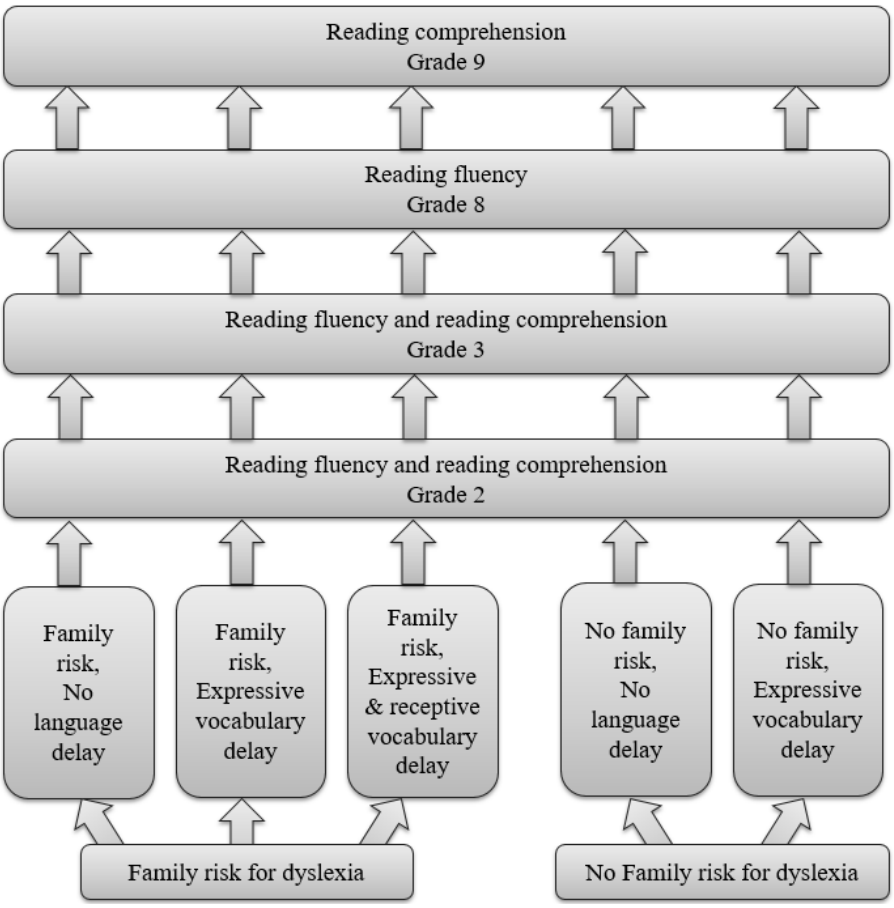

Figure 1 Design of Study I

Overall, the results showed that early vocabulary delay predicts reading comprehension, but not reading fluency development. The findings suggest that a pathway with reading comprehension difficulties later on is more probable when early expressive vocabulary delay is co-occurring with receptive vocabulary delay. Early expressive vocabulary delay alone was not a sufficient risk factor for the development of reading difficulties regardless of the presence of family risk for dyslexia. The findings showed that a delay in early vocabulary can lead to a persistent deficit with the specification that a deficit in both expressive and receptive vocabulary is a more powerful risk factor while the expressive vocabulary deficit alone can be alleviated over time. Only those with both expressive and receptive vocabulary delay and family risk for dyslexia had obvious reading comprehension difficulties that persisted into adolescence. Contrary to reading comprehension, reading fluency associated with family risk for dyslexia and not with the presence of early expressive vocabulary delay.

It may be concluded from this study that the reading trajectories from those with early vocabulary delays depend on the reading skill in question and the co-existence of other risk factors. Early expressive vocabulary delay represented a high-risk factor only for the development of reading comprehension and only when it co-occurred with other risk factors, in particular with receptive vocabulary delay and family risk for dyslexia. Early expressive vocabulary delay on the other hand, without the presence of receptive vocabulary delay, did not increase the risk for reading difficulties regardless of the presence of family risk for dyslexia. In addition, it seems the difficulties in reading comprehension displayed by those with both expressive and receptive vocabulary delay remain throughout primary and the secondary school, indicating that the combined deficit is a strong risk factor.

3.2 Study II

Longitudinal Stability of Reading Difficulties: Examining the Effects of Measurement Error, Cut-offs, and Buffer Zones in Identification

This study explored the stability of reading difficulties from Grades 2 to 6 and focused on the effects of measurement error and cut-off selection in the identification of reading difficulties and their stability with the use of simulations (see Figure 2). It addressed the methodological limitations of previous studies by (a) exploring difficulties in both reading fluency and reading comprehension, (b) examining a larger, and non-English sample, and (c) applying a model-based simulation analysis to explore the effects of measurement error and cut-offs in the identification of reading difficulties. There are three main advantages of the use of simulations in this study: (a) we can explore the effects of measurement error and the effects of single cut-off points on transitions; (b) we can analyze all possible combinations for reading fluency and reading comprehension difficulties between Grades 2 and 6 (persistent reading difficulties, late-emerging reading difficulties, resolving reading difficulties); and (c) the larger sample permits the identification of more groups that most likely would not have been identified in observed data because some of these groups might be small.

The focus of this study was on the development from the beginning to the end of primary school. In Finland, formal instruction of reading begins in Grade 1. However, because it would not have been feasible to assess reading comprehension accurately in Grade 1 due to the early stage of reading development, the first assessment time point was in Grade 2. Grade 6 was the second time-point and the end of primary school. The focus was on primary school and until Grade 6 because after Grade 6 children enter secondary school and are taught by subject teachers while in primary school they are taught by the classroom teacher. This study utilized data from The First Steps Study. Only children for whom data were available at both time points were included in the

analyses. Data for reading fluency and reading comprehension assessments from a total of 1,432 children were used.

The research questions of this study were as follows: (1) How stable are reading fluency and reading comprehension difficulties from Grade 2 to Grade 6? (2) What is the effect of measurement error on the estimation of reading difficulties stability over time from Grade 2 to Grade 6? (3) What is the effect of using single cut-offs compared to a buffer zone when examining the stability of reading difficulties over time from Grade 2 to Grade 6? (see Figure 2).

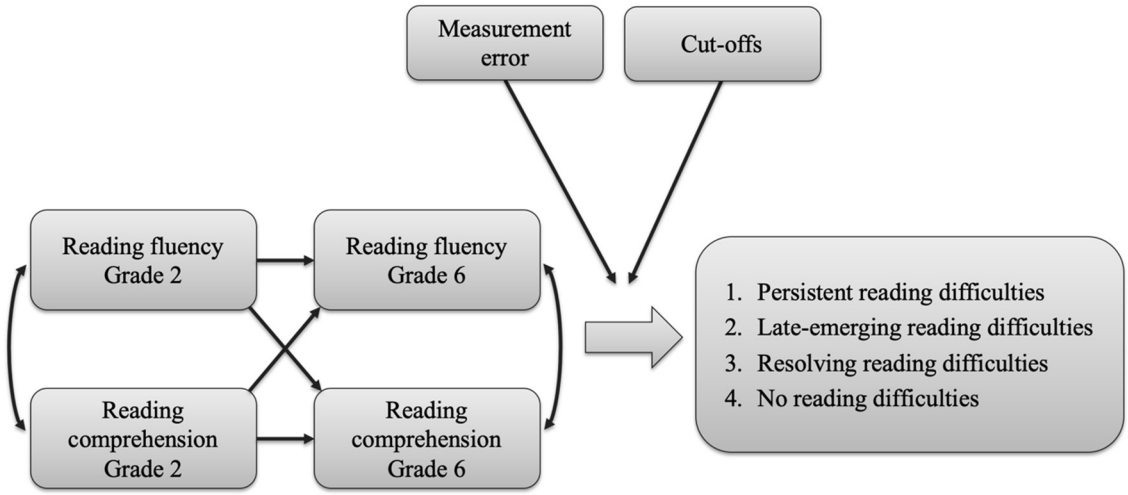

Figure 2 Design of Study II

In addition to the use of single cut-off points on observed data, a simulation approach was used in order to investigate the effect of measurement error on the stability of reading difficulties identification. The simulation analysis that explored the effects of measurement error on the identification of the groups with reading difficulties was based on a structural equation model which represented the development of reading fluency and reading comprehension from Grade 2 to Grade 6. Based on the structural equation model we simulated two datasets with 200,000 cases each: one using parameters related to latent factors, hence corresponding to true scores without measurement error; and the other using all the parameters in the model to produce data that also include the measurement error. Reading difficulties were identified as the lowest 10% of the distribution of reading fluency and/or reading comprehension. In addition to the effects of measurement error, we were also interested in the effects of single cut-offs on the stability of reading difficulties identification. For this reason, we also used a simulation-based buffer zone. Similarly to the use of single cut-offs, reading difficulties were identified as the lowest 10% of the distribution of reading fluency and/or reading comprehension. However, the lowest 10%–25% range

was considered to represent borderline scores, and scores above the 25% were considered to indicate no reading difficulties.

Our results aligned with prior findings, showing that for some children reading difficulties are not stable and the unstable groups (late-emerging and the resolving) truly exist. In the study, 74%–77% of participants were typical readers across time, and each reading difficulty group comprised 7%–10% of the participants (according to the group and the model). Concerning the stability of reading difficulties over time, compared to other studies, we identified a lower prevalence of late-emerging cases and a higher one of resolving. This could be due to differences in the criteria used for the identification of reading difficulties, in the assessment ages of the children, or in orthography. However, our results also revealed a clear effect of measurement error in the identification of reading difficulties over time. The simulation without measurement error suggested larger proportions of the stable groups (persistent and no reading difficulties) and smaller proportions of late-emerging and resolving reading difficulties in comparison to the simulation with measurement error. This finding demonstrates how measurement error affects the evaluation of the longitudinal stability of reading difficulties identification. In addition, the results showed that the cut-off selection also affects the identification of reading difficulties. In our study, the use of a simulation-based buffer zone unveiled a more complicated picture than the one using single cut-offs. Although many children who seem to resolve their reading difficulty and pass the criterion in order to be identified as having reading difficulties, they often score just above the cut-off but are in fact still at the lower end of the skill distribution. However, even after controlling for measurement error and using buffer zones, reading difficulties were not stable across time for all children, and both the resolving and the late-emerging groups were identified.

It may be concluded that measurement error has an impact on the accurate identification of children with reading difficulties by causing misclassifications and that the simplicity of the single cut-off points can lead to faulty impressions concerning the instability of reading difficulties across time. Overall, though, the findings supported that the unstable groups remain, even if we control for measurement error and use the buffer zone. A simulation-based buffer zone can control both the effects of measurement error and the arbitrariness of single cut-off points and contribute to more accurate classification into groups with reading difficulties, particularly for those with scores close to the cut-off point.

3.3 Study III

Developmental Profiles of Reading Fluency and Reading Comprehension from Grades 1 to 9 and Their Early Identification.

Considering that reading fluency and reading comprehension are not stable across time, the question arises about which additional factors predict the

development of reading fluency and especially the development of reading comprehension, which seems to be less stable. The aim of the third study is to investigate the different developmental profiles of reading fluency and reading comprehension development in Grades 1 to 9 (age 7 to 15) and to predict profile membership with kindergarten measures (cognitive skills, parental factors, and gender). The cognitive skills assessed were (1) phonological awareness, (2) letter knowledge, (3) rapid automatized naming of objects (RAN), (4) number counting, (5) word reading, (6) vocabulary, and (7) listening comprehension. The parental factors included (1) mother's education level, (2) father's education level, and (3) family risk for reading difficulties. The research questions of the study were as follows: (1) What kind of profiles can be identified based on the development in reading fluency and reading comprehension skills from Grades 1 to 9? (2) To what extent do child-related (i.e., gender, cognitive skills) and parent-related factors (i.e., mother's education, father's education, reading difficulties) assessed in kindergarten predict membership in the identified profiles? (see Figure 3).

This study utilized data from The First Steps Study and approximately 2,000 children were followed from kindergarten to Grade 9. Children's cognitive skills were assessed in kindergarten during fall and/or spring and their reading fluency and reading comprehension skills were assessed in Grades 1, 2, 3, 4, 6, 7, and 9.

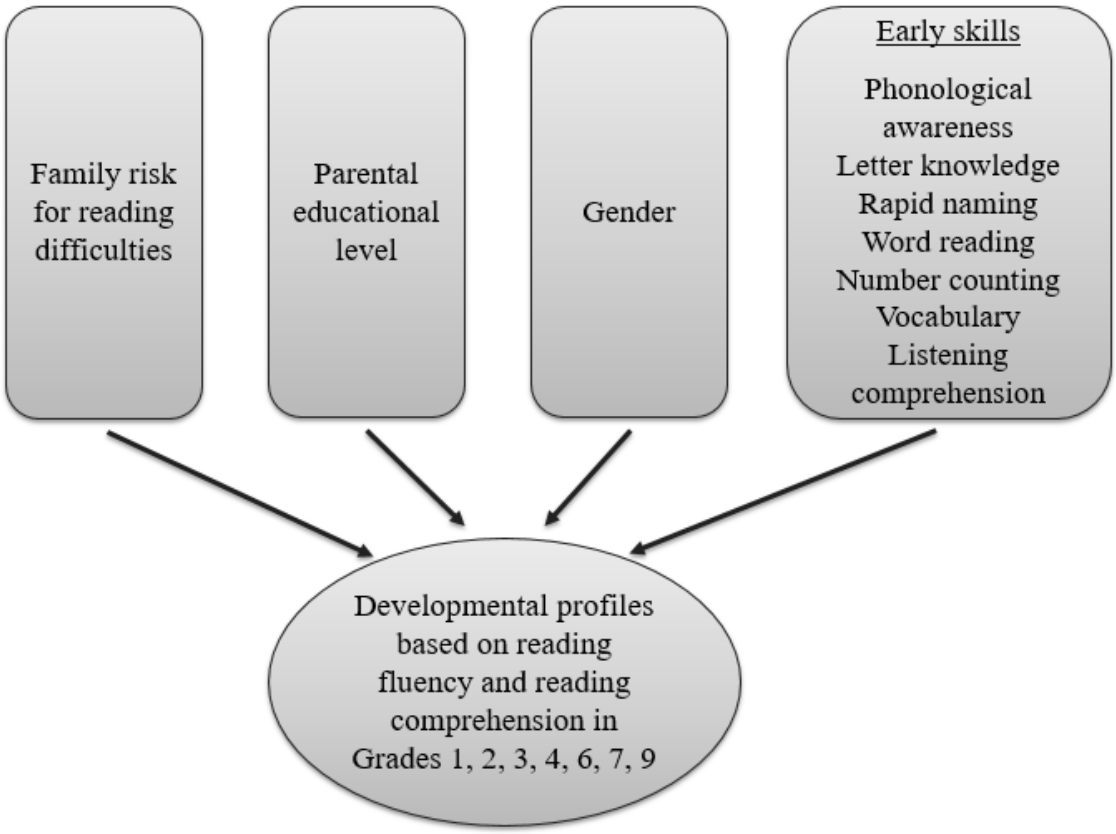

Figure 3 Design of Study III

The first step of the analysis was to identify the different profiles of reading fluency and reading comprehension development. The next step was to determine how cognitive skills, parental factors, and gender predicted membership in the identified profiles. For the identification of the different profiles of reading fluency and reading comprehension development we used a latent profile analysis, and four different profiles were identified: one with persistent reading difficulties across grades; one with early poor reading skills but with a resolving tendency; one with average reading skills; and one with good readers who started with very high reading fluency but scored average over time.

In line with earlier studies, we found an increased rate of reading difficulties among children with family risk for reading difficulties. In addition, similarly to previous studies, low performance on early cognitive factors increased the odds of developing reading difficulties. We also found increased vulnerability among boys; it was more likely to develop reading difficulties while it was less likely to resolve them compared to girls. Of special interest was the group that showed a resolving tendency to their reading difficulties: being a girl, good number counting skills, and vocabulary level seemed to be good marks for the resolution of reading difficulties.

4 DISCUSSION

The main objective of this dissertation was to examine the development of reading skills and the key factors for the early identification and resolution of reading difficulties. Overall, the results suggested that for most children with reading difficulties, the difficulties are persistent from Grades 1 to 9. There were, however, participants who manifested instability in their reading difficulties as well. In other words, reading difficulties in the early grades did not inevitably lead to reading difficulties throughout primary and lower secondary school for all children. Paths with late-emerging and resolving reading difficulties were also identified. Furthermore, the method of reading difficulties identification affected the conclusions that can be drawn about their stability over time. More specifically, the use of single cut-offs was likely to lead to faulty impressions, such as the distinctness of the reading difficulty groups. The use of different methodologies such as person- versus variable-oriented approaches may yield differential stability estimates. On the one hand, the use of latent profile analysis (a person-oriented approach) revealed higher stability in reading skills compared to the use of cut-offs on continuous variables (a variable-oriented approach). On the other hand, the examination of the longitudinal stability of reading difficulties with the use of cut-offs on continuous variables, which is a common practice, showed that the use of measures with measurement error and the use of single cut-offs affect the longitudinal stability of reading difficulties identification across two time points by causing misclassifications. Where and how we set a cut-off point matters and has ethical implications. The results of this dissertation show that follow-up of children's reading skills beyond the early grades is needed because it can reveal changes in children's reading status.

The second aim of this thesis, the early identification of the differential pathways in reading development, suggested multiple risk and protective factors. The odds of developing reading difficulties were increased by family risk for reading difficulties, being a boy, and lower performance in kindergarten-age cognitive skills (slow RAN, difficulty in reading easy words, and low scores in phonological skills, letter knowledge, number counting, and vocabulary). Factors such as being a girl as well as good number counting and vocabulary skills in

kindergarten seem to be good markers for the resolution of reading difficulties. Even though early cognitive skills were strongly related with reading difficulties later on, a profile with risk factors did not inevitably lead to reading difficulties. Family risk for reading difficulties was associated with difficulties in reading and particularly in reading fluency.

4.1 Stability of reading difficulties

Studies II and III examined the development of reading fluency and reading comprehension from Grades 1 to 9. Both studies showed that most of the children with reading difficulties in the early grades had difficulties across grades. However, some children showed instability in their reading difficulties. Study I suggested that there are stable reading comprehension difficulties across Grades 2 to 9 among children with language difficulties and familial risk for dyslexia.

The instability of reading difficulties across time is in accordance with prior studies (Etmanskie et al., 2016; Leach et al., 2003; Lipka et al., 2006; Torppa et al., 2015; van Viersen, 2019). These findings add to the literature by addressing their methodological shortcomings: small samples, use of single cut-offs and inclusion of measurement error. Due to the methodological issues there was severe uncertainty if these groups really exist or if they are the result of the effect of measurement error rather than the result of true change. Their conclusions may have been biased by false-positive and false-negative cases because of the inclusion of measurement error in their analyses, which can cause misclassifications. Only one study (Catts et al., 2012) had shown that the late-emerging and the resolving groups existed by using a type of latent class analysis with latent variables, and thus by not including measurement error in the analysis, but it examined English-speaking children, and although all the different groups of the late-emerging and the persistent reading difficulties had been examined, this was not the case for the resolving group.

Study II confirmed the hypothesis for heterogeneous developmental groups (Catts et al., 2012; Etmanskie et al., 2016; Leach et al., 2003; Torppa et al., 2015; van Viersen, 2019), showing that the instability is the result of true change and not only due to the effect of measurement error. The unstable groups could be identified even when the effect of measurement error was controlled and late-emerging and resolving groups were identified for difficulties in reading fluency, in reading comprehension, and in both skills. The findings of this study also revealed the disassociation between reading fluency and reading comprehension difficulties. For example, it appeared that at least average reading comprehension could be achieved even for children with difficulties in reading fluency. This corroborates previous studies that have identified different groups of reading difficulties: only reading fluency difficulties, only reading comprehension difficulties, both reading fluency and reading comprehension difficulties, and no difficulties (e.g., Catts et al., 2006, 2012; Hjetland et al., 2019; Nation et al., 2004, 2010; Snowling et al., 2009, as cited in Nation, 2019, p.61; Stothard et al., 2010, as

cited in Nation, 2019, p.61; Torppa et al., 2007). The findings also provided validation for the presuppositions of the SVR model (Gough & Tunmer, 1986; Hoover & Gough, 1990). Study II further showed through the comparison of two simulations (with and without measurement error) that reliance on methods which do not control for measurement error is prone to misclassifications and affects both the longitudinal stability of reading difficulties and the accurate identification of children with reading difficulties.

In Study III, four groups of reading fluency and reading comprehension development were identified using latent profile analysis: one with persistent reading difficulties across grades (7.8%), one with early poor reading skills with a resolving tendency (40%), one with average reading skills (42.3%), and one with good reading skills (9.9%). The identification of the resolving group, thus, gave partial support for the heterogeneous developmental groups hypothesis. It should be mentioned though that this study focused on the similarities and differences between individuals in reading development across time and on the identification of heterogeneous groups of individuals and not on the identification of reading difficulties. In addition, it is likely that not all those belonging to the resolving group in particular would be identified in school as having reading difficulties. The findings of Study III were also not fully compatible with the SVR model, because a group of poor comprehenders was not identified. The group with the persistent reading difficulties demonstrated difficulties in both reading fluency and reading comprehension skills during the whole nine-year period, from Grades 1 to 9. By the end of Grade 9, their reading fluency and reading comprehension continued to be approximately 1 SD below the scores of the other three groups. On the other hand, the children belonging to the group with the resolving tendency started at the same level as those with persistent reading difficulties but they showed steady progress in their reading skills across grades. By the end of Grade 9, the final assessment time point of this study, they had similar scores in comparison to the two groups with no reading difficulties. As Study II – which utilized the same follow-up data – identified a poor comprehenders group, it is possible that the latent profile analysis did not have enough statistical power to identify the small group of poor comprehenders. It should be noted, though, that this finding is in accordance with recent studies showing that children with poor decoding skills tend to have poor reading comprehension skills as well (Ferrer et al., 2015; Nation et al., 2019).

The findings of Studies II and III, even though they both come from the same longitudinal study, seemed at first to be in contrast regarding stability. The findings of Study III suggested mostly stability in reading skill development across Grades 1 to 9 while the findings of Study II give the impression that many changes in the reading difficulty status can happen during primary school from Grades 2 to 6. The findings from Study II thus seemed to be compatible with both the hypothesis for heterogeneous developmental groups and the SVR model. The results from Study III, however, were not fully compatible with either the hypothesis for heterogeneous developmental groups or the SVR model. Can the findings from these two studies be integrated even if they seem to be in contrast?

Part of the inconsistencies observed in the results can be attributed to the methodological differences of the studies. The aim of Study III was to identify subgroups of participants that share similar developmental pathways across time. Thus, in Study III, reading groups were not identified based on divergence from average performance. In Study II, however, the focus was on the identification of reading difficulties, i.e. the interest was in the lowest 10% of the reading skill distribution, and the stability of the development of children belonging to these subgroups at two time-points. Consequently, the groups with poorer performance in reading identified in Study III were not nearly as severe as in Study II. Thus, the results may suggest that reading difficulty stability is not exactly the same at the different places of the skill distribution. However, it is also possible that the use of simulated data of 200,000 cases in Study II allowed the identification of more groups that we were not able to identify in the observed data because some of these groups were small.

Another difference between Study II and III concerned the identification of late-emerging reading difficulties group (identified only in Study II). It is plausible that the relatively small group of late-emerging reading difficulties could not be identified in Study III using a latent profile analysis among the approximately 2,000 children. The developmental pathway of those with late-emerging reading difficulties was possibly not distinct enough from those of the poor readers for the latent profile analysis to pick it up. It is possible that within a bigger number of latent profiles, a group with late-emerging reading difficulties might have emerged. The number though of the optimal number of the profiles identified in Study III was based on specific criteria (Akaike, 1987; Bauer & Curran, 2003; Celeux & Soromenho, 1996; Lo et al., 2001; Muthén, 2003; Schwarz, 1978; Sclove, 1987; Vuong, 1989) that suggested that the four profiles described best the sample of the children and the different subgroups that could be identified. In Study II the prevalence of the group with late-emerging reading difficulties was lower (7%) than in previous studies. For instance, Catts et al. (2012) reported that 13% of the children had late-emerging reading difficulties. It is possible that late-emerging reading difficulties is a more prevalent phenomenon in the English context than in the transparent Finnish context. A previous study conducted among Finnish-speaking children, but in a different sample, (Torppa et al., 2015) also reported that there is a small group (10%) of children with late-emerging reading difficulties.

Despite the differences described above, both Study II and Study III identified a group with persistent reading difficulties. The reading performance of the children belonging to these groups was low from the beginning and deviated from the other groups even slightly more from Grade 4 onwards. That is, their reading fluency developed more slowly than among other groups. They also had a drastic drop in Grade 3 reading comprehension scores. Grades 3 and 4 seemed to be an important phase in reading development pathways and seemed to be in line with the "fourth-grade slump" that has been suggested (Chall & Jacobs, 2003). At around that grade level, the basic decoding skills have typically been acquired, and the focus of reading instruction also changes.

During the early grades, the focus is more on supporting children to become fluent readers while during the later grades reading is a tool for learning, the texts are more demanding, and they require more advanced vocabulary (Chall & Jacobs, 2003). The children belonging to the group with persistent reading difficulties might need more time to adapt to this change considering that texts include more complex words that require more advanced decoding skills. Thus, the Grade 3 slump in reading comprehension may reflect the fact that the text at hand was still overly difficult considering their dysfluent and struggling reading ability.

Both Study II and Study III found a group with resolving pathway. This finding is in accordance with previous studies (Catts et al., 2012; Torppa et al., 2015) but the proportions found were higher than previously. In Study II, 6.50% of the children resolved their reading difficulties whereas in Study III, 40% of the children belonged to the group with the resolving tendency. In Catts et al. (2012) study, only 1.90% of the children had resolving reading difficulties whereas in Torppa et al. (2015) 8.24% of the children had resolving reading difficulties. The prevalence of the group with resolving reading difficulties found in Study III is so high because the study did not focus on the identification of reading difficulties but on the differential developmental pathways as explained above. The children belonging to the resolving group in Study III started, however, in Grade 1 at the same level as those with persistent reading difficulties but by the end of Grade 9 they had scores only slightly below average both in reading fluency and in reading comprehension. By the end of Grade 9, their difference compared to those belonging to the group with persistent reading difficulties was approximately 1 SD both in reading fluency and in reading comprehension. The difference between these two groups is in the developmental trajectory they followed. The resolving group steadily became faster and faster in reading over time and almost caught up the groups with no reading fluency difficulties. In reading comprehension, this group manifested a very rapid development until Grade 3 and then their pathway stabilized to average. In Grade 9, they had almost equal scores with the groups with no reading comprehension difficulties.

It is possible that differences in the criteria used for the identification of reading difficulties, in the assessment ages, or in orthography, partially explain the differences in the larger prevalence of the resolving pathways found in Studies II and III and in the lower prevalence of late-emerging difficulties found in Study II compared to the findings of Catts et al. (2012). The study conducted by Catts et al. (2012) was the only previous study that assessed both word reading and reading comprehension, and in the analysis, they used multiple indicators for each reading class and thus their findings were less affected by measurement error than the other previous studies. Learning to read accurately in English is challenging because it is an opaque orthography. In transparent orthographies, such as in Finnish, there is one-on-one correspondence between graphemes and phonemes (Aro, 2017; Lyytinen et al., 2015; Seymour, Aro, & Erskine, 2003) and the acquisition of high accuracy is a relatively fast process. In transparent orthographies children typically read accurately after one year of formal

instruction of reading (Landerl & Wimmer, 2008; Lerkkanen et al., 2004; Soodla et al., 2015). By the end of Grade 2, the first assessment time point in Catts et al. (2012) as well as in Study II, most of the Finnish-speaking children are fluent readers and have good command of reading comprehension skills (Lerkkanen et al., 2010). It is thus likely that the transparency of the orthography may explain the fewer cases of the late-emerging reading difficulties identified in Study II or the unidentified group in Study III. It is also possible that the combined effects of systematic phonics instruction, special education, and the transparency of Finnish orthography could enhance the development of a resolving pathway more often than in less transparent orthographies.

One might argue that also the difference in the time gap between the identification of the early and late reading difficulties might make a difference and that the longer gap could as well increase the prevalence of the cases with unstable reading difficulties. In Catts et al. (2012) study, children were assessed in Grades 2, 4, 8, and 10 whereas in Study II children were assessed in Grades 2 and 6. However, if the longer gap would explain the differential prevalence of the unstable groups, we should have also found a larger group of late-emerging cases. On the contrary, fewer late-emerging cases were found than in the study by Catts et al.

Despite the issues mentioned above on the comparability of the findings of Study II and Study III and their comparability with previous studies examining the stability of reading development across time, four conclusions can be drawn. First, although for many children difficulties are stable over time and continue across school years, there are children who present instability in their reading difficulties, in particular children who manage to resolve them and end up showing close to average reading skills. Second, the use of measures with measurement error and the use of single cut-offs affect the longitudinal stability of reading difficulties identification across two time points. Third, the use of a single cut-off can lead to faulty impressions, such as the distinctness of groups with reading difficulties. Researchers should be aware of this and it should also be communicated to practitioners. Where and how we set a cut-off point matters and has ethical implications. It affects not only who will be identified as having reading difficulties but also who will receive extra support. This is particularly significant in education systems where access to extra support depends on an official diagnosis and the classification into a learning difficulty category. Fourth, follow-up of children's reading skills beyond the early grades is needed because it can reveal notable changes in children's reading status.

4.2 Early predictors of reading difficulties

The second aim of the dissertation focused on the examination of the factors for the early identification of the different reading pathways. The main aim was to increase understanding of how early reading-related cognitive and language skills, parental factors (family risk and parental education level), and gender

predict the development of reading skills. Data were drawn from two different longitudinal datasets: the JLD study (Study I) and The First Steps Study (Study II and III). The JLD as a family risk study offered very thorough assessment of family risk and the development of early language and literacy. The First Steps Study provided a large population-based sample and included the assessment of a broad range of cognitive and language skills in kindergarten and academic skills in primary and secondary school grades. Most of the well-known key skills that predict reading skills development were included in the analysis: phonological awareness, RAN, letter knowledge, vocabulary, and listening comprehension. In addition, the effects of number counting (Study III) – a skill that is not usually included in reading studies – were examined, along with the effects of family risk for reading difficulties (Studies I and III), parental education (Study III), and gender (Studies II and III) in order to identify possible early predictors of reading difficulties.

4.2.1 Cognitive skills

In the identification of the different pathways of reading development, specific cognitive skills can provide critical information on ways to identify, early on, the children who will have differential developmental pathways. In line with previous studies, Study III found strong predictive links between cognitive skills and difficulties in reading fluency and/or reading comprehension (Araújo et al., 2015; Caravolas et al., 2019; Clayton et al., 2019; Landerl & Wimmer, 2008; Melby-Lervåg et al., 2012; Petscher et al., 2018; Puolakanaho et al., 2008; Snowling et al., 2019b; Torppa et al., 2016). Higher levels of cognitive skills assessed in kindergarten-age increased the probability of not developing reading difficulties later on.

Among the cognitive predictors, word reading, RAN, number counting, and phonological awareness were the most significant factors for the development of early reading difficulties. Both the group with persistent reading difficulties and the group with the resolving tendency in Study III had significantly lower scores than the two groups with no reading difficulties. This finding is in accordance with the vast literature suggesting that these skills are the key predictors of reading accuracy and fluency (Caravolas et al., 2019; Clayton et al., 2019; Koponen et al., 2013, 2016; Landerl et al., 2019; Melby-Lervåg et al., 2012; Puolakanaho et al., 2008; Snowling et al., 2019; van Viersen et al., 2018; Ziegler et al., 2010) and contribute to the early reading difficulties these children experienced. Considering that at least average reading fluency skills constitute a prerequisite for reading comprehension (Florit & Cain, 2011; García & Cain, 2014; Torppa et al., 2016), it was shown, as expected, that children with low fluency skills are likely to have reading comprehension difficulties during the early grades. Those with persistent reading difficulties during primary and secondary school were found to have low scores in all included kindergarten-age cognitive skills except from listening comprehension in comparison to the two groups with no reading difficulties. Children belonging to the group with the resolving tendency had lower scores in word reading, RAN, number counting, letter

knowledge, and phonological awareness in comparison to those with average or good reading skills. This finding could suggest a broader deficit in cognitive skills for those with persistent reading difficulties compared to those with the resolving tendency. On the other hand, those with the resolving tendency were likely to have difficulties mainly in decoding and not in early language skills.

An interesting finding from Study III was the identification of significant kindergarten indicators for membership in the group with tendency for the resolution of reading difficulties (difference between those with persistent reading difficulties and those with the resolving tendency). Good number counting and vocabulary skills in kindergarten as well as being a girl seemed to increase the probabilities for a resolving reading difficulties pathway. The underlying link between number counting and reading in the early phases of reading development could be the one-by-one processing which is essential for both skills. Although not much is known about number counting, as it is rarely included in reading studies, its contribution to reading fluency skills has also been shown previously even after controlling phonological awareness, verbal short-term memory (Koponen et al., 2013, 2016), vocabulary, working memory, number concepts, and maternal education (Koponen et al., 2016). It has even been reported to be a stronger predictor than RAN (Koponen et al., 2016). This is noteworthy because in transparent orthographies such as Finnish, RAN would be expected to be the strongest predictor of reading fluency (Georgiou et al., 2008; Landerl et al., 2019; Moll et al., 2014). Those with the resolving tendency would, thus, be expected to have better rapid naming skills than those with persistent reading difficulties and their difficulties would be more in decoding and phonological awareness. However, in line with earlier findings (Koponen et al., 2016), the study indicated that number counting could be a stronger predictor than RAN. Even though both RAN and number counting require serial processing, each requires different knowledge. In contrast to the RAN task in which the child needs to name visual stimuli as fast as possible, in the number counting task, a child needs to hold information in their memory in order to find correct responses.

Limited early vocabulary, on the other hand, can limit the words available for recognition, which, in turn, can affect the ability to comprehend text. Strong predictive effects of very early vocabulary on reading development were identified in Study I, which also showed a differentiation between these effects on reading fluency and reading comprehension. Children with delays in early expressive and receptive vocabulary developed difficulties in reading comprehension but not in reading fluency in Grades 2 to 9. This finding is in accordance with earlier studies suggesting an association between early vocabulary and later reading comprehension skills (Duff et al., 2015; Lyytinen et al., 2005; Rescorla, 2005; Rescorla, 2009) but the present study extends the findings to a 13-year follow-up. The findings showed that reading comprehension was strongly predicted not by a delay in onset of expressive vocabulary, but by a delay in both early expressive and receptive vocabulary (see also Eklund et al., 2018). Early expressive vocabulary delay thus seems to

alleviate in time and not to be a sufficient predictor of school-age reading skills. Study III further showed that an assessment of vocabulary even using a short-form scale in kindergarten can also provide important information on pathways of reading development during primary and secondary school.

Study I was limited though by not having a group with both expressive and receptive vocabulary delay but no family risk. In the present sample, children with both expressive and receptive vocabulary delay also had family risk, making it difficult to differentiate the effects of family risk and the effects of receptive vocabulary delay. Nonetheless, it seems that family risk does not cause greater risk for reading comprehension difficulties in the present sample as the two other groups with family risk (either with no vocabulary delay or with only expressive vocabulary delay) manifested average reading comprehension skills. Moreover, the greater prevalence of the combination of the expressive and the receptive vocabulary delay among those with family risk indicates an association between family risk for reading difficulties and a combined vocabulary delay. The combination of the family risk for reading difficulties and early delays in expressive and receptive vocabulary was also observed in a recent study among English-speaking children (Duff et al., 2015).

Contrary to reading comprehension, reading fluency was not affected by delays in early expressive vocabulary, which is also in line with Rescorla's (2005) findings. This finding seems to oppose the lexical restructuring hypothesis (e.g., Walley et al., 2003), which proposes a link between early vocabulary and reading fluency mediated via phonological skills. However, although a previous study in the same sample showed that early vocabulary is associated with precocious reading ability in kindergarten through phonological skills (Torppa et al., 2007), in the present study the focus was on the later grades of primary and secondary school. At this later stage of reading development and in the context of a transparent orthography, phonological skills may not be a powerful predictor of later reading fluency (see e.g. Aarnoutse et al., 2005; Georgiou et al., 2012; Landerl et al., 2019), hence the mediated effect from vocabulary is not significant.

4.2.2 Family risk

Evidence showing that family risk for reading difficulties significantly affects children's reading skills and the development of reading difficulties was found also in the present dissertation. The predictive value of family risk was identified both in Study I and III. In Study III both groups with early reading difficulties (persistent and resolving) had increased rates of children with family risk for reading difficulties. In Study I, it could be further specified that the effect of family risk due to parental dyslexia was specifically on reading fluency but not on reading comprehension. This is expected based on the previous studies on the effects of family risk on reading difficulties (e.g. Snowling & Melby-Lervåg, 2016). Children with family risk had lower scores in reading fluency in Grades 2, 3, and 8 compared to those with no family risk regardless of the presence of early vocabulary delay. The effect sizes for the group differences were of medium size supporting the clear impact of family risk in developmental dyslexia (Lyytinen

et al., 2015; Snowling et al., 2003; Snowling & Melby-Lervåg, 2016; van Bergen et al., 2012).

Although in Study III family risk seemed to be a good predictor of developing average or good reading skills, it did not seem to be an indicator for the resolution of reading difficulties and increased rates of family risk cases were identified also in the group with the resolving tendency. Except from number counting, vocabulary, and gender, no other factor was found to explain the resolving tendency. The measures used though for the familial effects were limited to parental reported reading difficulties and parental education. Therefore, the effects of, for example, other environmental factors at home or at school which facilitate the resolution of reading difficulties and affect reading development (e.g., book reading, access to special education, teacher effects; see Mol & Bus, 2011; Senechal & Le Fevre 2002; Torppa et al., 2019) remain unexplored in this dissertation.

The identification of family risk differed in Studies I and III. In Study I, a careful, comprehensive assessment through questionnaire, interview, and individual testing took place while only a simple questionnaire was used in Study III. Despite the weaknesses on the assessment of family risk in Study III, family risk was still a significant predictor of children's reading development. Although parental self-reports of reading difficulties are not as accurate as formal assessments for the determination of family risk for reading difficulties, it seems that self-reports capture some of the variance and yield a measure that could be used in the prediction of children's skills. Our results on the significant effects of parental reading difficulties on children's reading skills were in line with a previous study that has also used parental self–reports for the assessment of family risk (Esmaeeli et al., 2019; see also Snowling et al., 2012). Consequently, because it is a more feasible tool for practitioners in the schools, it could be utilized for the early prediction of children's reading difficulties and as a means for these children to obtain early access to extra support.

4.2.3 Gender

Gender effects were investigated in Studies II and III. In Study III, gender was found to be a significant indicator for the resolution of reading difficulties over time. Gender effects on the resolution of reading difficulties were also reported in a previous Finnish study conducted with a different sample (Torppa et al., 2015). Both studies reported that it is more likely for girls to manifest a resolving trajectory in their reading difficulties than it is for boys to do so. In Study II, boys were also found to be over-represented in the groups with late-emerging and persistent reading difficulties.

Overall, both studies found an increased vulnerability for reading difficulties in boys. This finding is in line with previous studies suggesting that girls outperform boys (Berninger et al., 2008; Clinton et al., 2014; Hawke et al., 2009; Quinn & Wagner, 2015; Rajchert et al., 2014; Rutter et al., 2004; Stoet & Geary, 2013) and that more boys than girls are identified with reading difficulties (Hawke et al., 2007; Quinn & Wagner, 2015; Stoet & Geary, 2013).

4.2.4 Early predictors in sum

Despite the differences in the samples and the limitations the studies had, four conclusions can be drawn about the early prediction of reading difficulties. First, family risk for reading difficulties, being a boy, and lower performance in cognitive skills (slow RAN, low scores in phonological skills, letter knowledge, number counting, and vocabulary, and difficulty in reading easy words) in kindergarten increase the odds of developing reading difficulties. Second, reading difficulties in Grade 1 do not inevitably lead to reading difficulties throughout primary and secondary school. Factors such as being a girl, and good number counting and vocabulary skills in kindergarten seem to be good markers for the resolution of reading difficulties. Third, even though early cognitive skills were strongly related with reading difficulties later on, a profile with risk factors does not inevitably lead to reading difficulties. For example, children with no family risk for dyslexia but with only early expressive vocabulary delay had average reading skills. Fourth, family risk for reading difficulties is associated with difficulties in reading and particularly in reading fluency.

4.3 Methodological limitations of the studies

In general, the studies composing the present dissertation have some limitations, that need to be taken into account when drawing conclusions. Some of them are specific to each study while others are more general and arise due to the methodology chosen. Study I had a relatively small sample that came from a prospective family risk study. The use of a larger sample would have provided greater statistical power. In addition, a larger sample may have helped in the identification of a group of children with both expressive and receptive vocabulary delay and no family risk for dyslexia. This group would have been informative in the examination of the effects of family risk and the effects of early vocabulary delay on later reading skills development. Moreover, inclusion of vocabulary measures at school age, would have allowed to examine whether the identified differences among the groups on vocabulary measures continued and co-occurred with reading difficulties in adolescence. Furthermore, the statistical analysis used in this study is accompanied by the problems that were discussed in Study II, that is, use of measures with measurement error and the use of arbitrary cut-offs. It is thus likely that the identified groups include some false-positive and/or false-negative cases.

Study II and Study III utilized data from The First Steps Study in which reading fluency was assessed with three measures at each grade, whereas reading comprehension with only one measure. In Study II we calculated the correction of attenuation in order to set measurement error also for reading comprehension, but having more measures would have strengthened the model. In addition, no data were collected for Grades 5 and 8 in the follow-up which formed a gap preventing modelling of reading fluency and reading

comprehension skills development across all grades up to Grade 9. However, longitudinal studies like the JLD and The First Steps Study, with such long-term follow-up periods are rare and provide unique possibilities despite some compromises that were needed. Another limitation is related with the listening comprehension measure used in Study III. The reliability of this measure is quite low (.31) which makes the measure weak and the results should be interpreted with caution.

Another limitation concerns the family risk variable used in Study III. Although there were data for both mothers' and fathers' reading difficulties, there were many missing values for the fathers' self-reports. For this reason, a new variable was used for family risk in which either the mother or the father reported reading difficulties. If the two separate variables for mothers' and fathers' reading difficulties had been used, more information would have been available on whose difficulties affect children's reading skills more. The use of the two separate variables for mothers' and fathers' reading difficulties could also have provided an opportunity to examine whether the cognitive skills of those children with two parents with reading difficulties differ from those who have only one parent with reading difficulties. In addition, the family risk variables were based on self-reports using single items. Although this, as described above, can be used as an everyday tool, it raises questions on what adults conceive of as reading difficulties. When adults are asked whether they have reading difficulties, do they focus more on reading fluency or on reading comprehension? The use of a more detailed parental self-report or individual assessment of parents' skills would have provided a better assessment of parental reading difficulties and of what they view as reading difficulties.

4.4 Concluding remarks and future directions

The overall aim of the current dissertation was to examine the development of reading skills over time (ages 2 to 16) and investigate the key factors for the early identification and resolution of reading difficulties. Moreover, the dissertation aimed to increase understanding of how family risk, cognitive skills, and environmental effects affect reading development and reading difficulties. Successful reading is essential in order for an individual to fully participate in today's society. From educational and professional settings to a variety of cultural and social activities – all of these require the assimilation of information from written texts. It is thus essential to build accurate models that show how reading skills are developed so that we can establish evidence-based curricula and support systems for children and youth facing difficulties in reading. However, because reading is a complex skill relying on many components and processes, the specification of accurate models is challenging.

The results of this dissertation indicate that reading difficulties are not persistent across grades. For some children, their reading difficulties remain persistent across grades. Yet there are some children who manage to resolve their

early reading difficulties and others who start school without having reading difficulties but manifest them later on. The methodology that is being used by researchers highly affects the estimation of the stability of reading skills development and the stability of reading difficulties identification. The use of cut-off points is likely to lead to uncertainties in research findings because of measurement error. In addition, the use of cut-off points would be justified if we did not have a normal skill distribution, but this is not the case in reading achievement (Francis et al., 2005). The use of single cut-offs could contribute to false or biased estimations on the prevalence of instability of reading difficulties if those belonging to the unstable groups are scoring just above or just below the cut-off point. This can lead to faulty impressions about how distinct the groups are as well as about the risk of children for developing reading difficulties. Researchers need to be aware of these implications and they also need to communicate them in practice. Where and how we set a cut-off point matters and has ethical implications. It affects not only who will be identified with reading difficulties but also who will receive extra support. However, because cut-off points are a practical everyday tool for the identification of children with reading difficulties, we need to find ways to deal with the problems that accompany them. Traditional statistical tools do not seem to be well suited. Better ways and new approaches to model individual differences across time are required.

One approach could be the use of buffer zones that control the arbitrariness of single cut-off points. Another approach is the use of methods that control the effects of measurement error and do not use cut-off points. Such ways can lead to more reliable examination of the developmental trajectories of reading skills and to a more accurate identification of children with reading difficulties. This is particularly significant for education systems in which the access to special needs interventions and extra support depends on the classification into a learning difficulty category and an official diagnosis is needed. Both approaches were employed in the current dissertation and their use suggested that reading skills are not always stable across grades in primary and secondary school. In practice, skill assessments always include measurement error, so the reliability of the assessment needs to be increased via multiple tasks and repeated assessments. Consequently, children's reading skills need to be followed carefully from early on and also beyond the early grades because changes in their reading difficulty status could occur at any time point. Only continuous follow-ups can identify the children who may experience reading difficulties later on during their academic path. On the other hand, there is a need for a revision of the practices of diagnosing reading difficulties as well as the persistence of the diagnoses of reading difficulties. This would be especially beneficial for children with late-emerging reading difficulties who are deprived of interventions or support systems, and for children with resolving reading difficulties who must carry a label that is not accurate for their entire lives.

The findings of this dissertation showed that groups with different developmental trajectories of reading skills can be identified and they show differences in their early reading-related cognitive skills. In the current

dissertation, I was unable to identify one single factor that is sufficient on its own to predict children's later reading skill. This is in accordance with the multiple deficit model (Pennington, 2006) and its extension, the intergenerational multiple deficit model (van Bergen et al., 2014b). Children belonging to the groups with reading difficulties showed multiple deficits in several cognitive skills as well as differences in parental factors. Low scores in early cognitive skills and the presence of family risk could lead to later reading difficulties. Both the group with persistent reading difficulties and the group with the resolving tendency in Study III had poorer scores in word reading, RAN, number counting, and phonological awareness compared to those with no reading difficulties. The group of children with persistent reading difficulties across grades also had lower scores in letter knowledge and vocabulary compared to the groups with no reading difficulties. However, as Study I showed, a profile with high risk (e.g., family risk for dyslexia and early expressive vocabulary delay) does not inevitably lead to reading difficulties. The effect on reading comprehension was shown only if receptive vocabulary was also impaired. In addition, the children belonging to the group with the resolving tendency demonstrated significant difficulties in multiple cognitive skills, but at the same time they had certain strengths in early cognitive skills in comparison to the children belonging to the group with persistent reading difficulties. Specifically, the probabilities for the resolution of reading difficulties increased with higher scores in number counting and vocabulary in kindergarten. These skills may have acted as protective factors, or indicators of skills that predict better outcomes in the long run in reading. Furthermore, gender seems to affect the trajectories of reading skills development, with girls being more likely to resolve their reading difficulties. These factors that seem to protect against the persistence of reading difficulties and help in their resolution, could provide the ground for planning support systems based on them.

In addition, the results of this study indicated in line with prior literature that family risk is a factor that can provide very valuable information on the development of children's reading skills. Although not all of those with family risk for reading difficulties develop reading difficulties themselves, it seems to increase predictability of children's reading difficulties. In everyday settings, it is not always possible to assess parents' reading skills for the accurate determination of family risk as in research settings. However, the results from Study III show that even simple parental self-reports can be used and reveal differences among those with and without reading difficulties. This kind of information could open the door for early support in pre-reading skills that have been found to facilitate later reading acquisition. It could also help the child to avoid the severe reading difficulties they would have faced without the early support or aid in resolving their reading difficulties. Further research is needed though for the factors that help the resolution of reading difficulties.

For future studies, two lines of development can be recognized: the need for large-scale longitudinal studies to identify the reasons for the individual difference in developmental pathways and the need for the development of new

methodological approaches. Most of the existing studies focus on the development of reading during the early grades. Much less is known about the reading skills development beyond the early grades and the reasons for the resolution of reading difficulties or for adolescents' poor reading skills. Better understanding is needed for both theory building and the development of systems that can be used to support children's reading skills. Large-scale longitudinal studies with repeated assessments at several time points and several measures can aid in more accurate mapping of the changes from grade to grade and the influence of reading-related cognitive and language factors, environmental factors, and factors that are related to motivation and task avoidance behavior. The use of the traditional statistical tools seems to create limitations for the examination of such issues. New methodological approaches to model differences across time are needed. One such example could be the use of the random-intercept cross-lagged models (Hamaker et al., 2015; Mund & Nestler, 2019), which allow the separation of within- and between-person effects and each individual fluctuates around their stable level. With the use of such models, we could examine the change in reading skills across time by taking the initial level out. Studies following children's reading skills and examining possible precursors are valuable, but a further step needs to be taken. More attention is needed on identifying those factors that affect reading skills during their development, e.g., the impact of instruction or access to extra support, cognitive skills that act as protective factors in reading difficulties, teacher's role, motivation, and the home literacy environment.

YHTEENVETO (FINNISH SUMMARY)

Lukutaidolla on nyky-yhteiskunnissa yhä kasvava merkitys, koska se auttaa meitä toimimaan oikein ja itsenäisesti. Yksi koulutuksen päätavoitteista onkin opettaa lapset lukemaan ja käyttämään lukutaitoaan oppimiseen. Lasten lukemisvaikeudet ovat vakava riskitekijä, joka vaikuttaa kielteisesti opintomenestykseen, myöhempään uraan, päivittäiseen toimintaan ja yhteiskunnalliseen osallisuuteen. Vaikka suomea pidetään hyvin läpinäkyvänä kielenä (Aro, 2017) ja lähes kaikki lapset osaavat lukea virheettömästi ensimmäisen luokan loppuun mennessä (Lerkkanen ym., 2004; Soddla ym., 2015), lukemisvaikeuksia havaitaan 5–20 %:lla lapsista (Lerkkanen ym., 2010). On yhä enemmän näyttöä siitä, että kaikki lasten lukemisvaikeudet eivät ole pysyviä, päinvastoin kuin yleensä oletetaan. Joillekin lapsille kehittyy lukemisvaikeuksia alakoulun kolmannen tai neljännen luokan jälkeen, vaikka heidän lukutaitonsa olisi aluksi ollut normaali. Toisilla taas on normaali lukutaito kolmannen tai neljännen luokan jälkeen huolimatta alaluokilla havaituista lukemisvaikeuksista (Catts ym., 2012; Etmanskie ym., 2016; Leach ym., 2003; Lipka ym., 2006; Torppa ym., 2015). Syitä nuorten heikolle lukutaidolle ja lukemisvaikeuksien korjaantumiselle ei vielä tiedetä. Sekä teoreettisesta että käytännön näkökulmasta on tärkeää tutkia lukemisen kehittymistä pitkällä aikavälillä. Samoin tulee tutkia tekijöitä, jotka auttavat tunnistamaan ajoissa riskit ja korjaamaan lukemisvaikeudet.

Tämän väitöstutkimuksen yleisenä tavoitteena on tarkastella lukutaidon kehittymistä pidemmän ajan kuluessa (ikävuosina 2–16) sekä lukemisvaikeuksien varhaiseen tunnistamiseen ja korjaamiseen liittyviä tekijöitä. Tutkimus pyrkii tunnistamaan tapoja, joilla eri-ikäisten lasten lukutaidon ja lukemisvaikeuksien kehittymistä voidaan ennakoida tarkastelemalla hyvin varhaisia taitoja. Erityisesti pyritään selvittämään, kuinka perheeseen liittyvät riskitekijät, kognitiiviset taidot ja ympäristötekijät vaikuttavat lukutaidon kehittymiseen ja lukemisvaikeuksiin. Väitöskirja koostuu kolmesta määrällisestä tutkimuksesta ja perustuu seuraavissa tutkimuksissa saatuihin tuloksiin: 'Lukivaikeuden pitkittäistutkimus syntymästä 23-vuotiaaksi' ja 'Alkuportaat'. *Tutkimuksessa I* keskityttiin hyvin varhaiseen ekspressiiviseen ja reseptiiviseen sanavarastoon, joita mitattiin 2- ja 2,5-vuotiailla lapsilla. Tarkoituksena oli selvittää, ennakoivatko perheeseen liittyvät riskit ja 2–2,5-vuotiaan ekspressiivinen ja reseptiivinen sanavarasto lukemissujuvuuden ja luetun ymmärtämisen kehitystä luokilla 2–9. *Tutkimus II* kartoitti lukemisvaikeuksien pysyvyyttä alakoulussa (luokilla 2–6) ja aiempien tutkimusten metodologisia ongelmia, jotka aiheuttivat huomattavaa epävarmuutta siitä, oliko epävakaita ryhmiä todella olemassa vai johtuivatko ne mittausvirheestä. *Tutkimuksessa III* pyrittiin selvittämään lukemissujuvuuden ja luetun ymmärtämisen kehitysprofiileja ensimmäiseltä luokalta yhdeksänteen luokkaan sekä kognitiivisten taitojen, sukupuolen, vanhempien koulutustason ja vanhempien lukemisvaikeuksien ennustavuusarvoa.

Väitöskirjan ensimmäistä tavoitetta eli lukemissujuvuuden ja luetun ymmärtämisen kehitystä koskevat tulokset osoittavat, että lukemisvaikeudet eivät aina jatku luokalta toiselle. Joillakin lapsilla on pysyviä lukemisvaikeuksia, mutta osa pystyy korjaamaan vaikeudet ja osalle taas ilmaantuu lukemisvaikeuksia vasta ylemmillä luokilla. Käytetyt menetelmät vaikuttavat suuresti lukutaidon kehityksen vakauteen sekä lukemisvaikeuksien tunnistamisen vakauteen. Tarkkojen raja-arvojen käyttö aiheuttaa todennäköisesti mittausvirheistä johtuvaa tutkimustulosten epävarmuutta. Raja-arvojen käyttö olisi perusteltua, jos taidot eivät noudattaisi normaalijakaumaa, mutta näin ei lukutaidon kohdalla ole (Francis ym., 2005). Yksittäisten raja-arvojen käyttö voisi johtaa vääriin tai yksipuolisiin arvioihin lukemisvaikeuksien epävakaisuuden yleisyydestä, jos epävakaisiin ryhmiin kuuluvat sijoittuisivat vain hieman raja-arvon ylä- tai alapuolelle. Tämä voi tuottaa vääriä tulkintoja ryhmien eroavuuksista sekä lasten lukemisvaikeusriskeistä. Tutkijoiden tulee tiedostaa nämä seuraukset ja myös tuoda ne julki. Sillä, minne ja miten raja-arvo asetetaan, on eettisiä seurauksia. Sen perusteella määritetään, kenellä on lukemisvaikeuksia, mutta myös se, kuka saa erityistukea. Lisäksi tulosten perusteella lasten lukutaitoa tulisi seurata myös alaluokkien jälkeen, sillä tilanne saattaa muuttua.

Toisena tavoitteena oli tutkia tekijöitä, joita voidaan käyttää lukutaidon erilaisten kehityspolkujen varhaisessa tunnistamisessa. Tulokset toivat esille ryhmiä, joiden lukutaidon kehityskaaret eroavat toisistaan ja joiden varhaisissa lukemiseen liittyvissä kognitiivisissa taidoissa on eroja. Emme löytäneet yhtä yksittäistä tekijää, joka ennustaisi lasten myöhempää lukutaitoa. Tämä tulos on yhtäpitävä "useiden riskitekijöiden mallin" (multiple deficit model, Pennington, 2006) ja sen jatkeen, "ylisukupolvisen useiden riskitekijöiden mallin" (intergenerational multiple deficit model, van Bergen ym., 2014b) kanssa. Lukemisvaikeuksia omaavien lasten ryhmissä esiintyi paljon kognitiivisia haasteita ja vanhempiin liittyviä eroja. Heikot varhaiset kognitiiviset taidot ja perheriskitekijä voivat johtaa myöhempiin lukemisvaikeuksiin. Sekä pysyvien lukemisvaikeuksien ryhmä että vaikeudet korjannut ryhmä tutkimuksessa III saivat heikommat tulokset sanojen lukemisessa, nopean sarjallisen nimeämisen testissä (RAN, rapid automatized naming), numerolaskuissa ja fonologisessa tietoisuudessa kuin lapset, joilla ei ollut lukemisvaikeuksia. Pysyvien vaikeuksien ryhmällä oli myös heikompi kirjaintuntemus ja sanavarasto kuin ryhmillä, joilla ei ollut lainkaan lukemisvaikeuksia. Riskiprofiili (esimerkiksi geneettinen dysleksiariski ja varhaisen ekspressiivisen sanavaraston viive) ei kuitenkaan väistämättä johda lukemisvaikeuksiin. Vaikka lukemisvaikeutensa korjaavan ryhmän lapsilla oli huomattavia vaikeuksia useissa kognitiivisissa taidoissa, heillä oli samanaikaisesti tiettyjä varhaisten kognitiivisten taitojen vahvuuksia verrattuna ryhmään, jonka lukemisvaikeudet olivat pysyviä. Erityisesti parempi numerolaskutaito ja sanavarasto esikoulussa lisäsivät lukemisvaikeuksien korjaantumisen todennäköisyyttä. Ne saattoivat toimia suojaavina tekijöinä tai osoituksena taidoista, jotka ennustavat parempia lukemistuloksia pitkällä aikavälillä. Sukupuoli näyttää myös vaikuttavan

lukutaidon kehityskaariin siten, että tyttöjen lukemisvaikeudet todennäköisemmin korjaantuvat. Vaikka sukupuolen merkitystä ei voida vähentää, tukijärjestelmien suunnittelussa voidaan huomioida tekijät, jotka näyttävät suojaavan pysyviltä lukemisvaikeuksilta ja korjaavan vaikeuksia. Tämän tutkimuksen tulokset osoittavat myös, että perheriskitekijä voi tuottaa hedelmällistä tietoa lasten lukutaidon kehityksestä. Kaikille ei kehity lukemisvaikeuksia perheriskistä huolimatta, mutta perheriski näyttää kuitenkin lisäävän lapsen lukemisvaikeuksien ennustettavuutta.

On tärkeää tutkia lasten lukutaidon kehitystä ja mahdollisia merkkejä tulevista vaikeuksista. Tutkimuksissa tulisi kuitenkin mennä vielä pidemmälle ja alkaa määrittää tekijöitä, jotka vaikuttavat lukutaitoon sen kehittymisen aikana. Tällaisia tekijöitä ovat opetus, erityistuen saatavuus, lukemisvaikeuksilta suojaavat kognitiiviset taidot, opettajan rooli, motivaatio ja kodin tekstitaitoympäristö.

REFERENCES

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

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332. https://doi.org/10.1007/BF02294359

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Association

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868–883. https://doi.org/10.1037/edu0000006

Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading. *European Journal of Psychology of Education, 16*(3), 363–383. doi: 10.1007/BF03173188

Aro, M. (2017). Learning to read Finnish. In L. T. W. Verhoeven & C. A. Perfetti (Eds.), *Reading acquisition across languages and writing systems: an international handbook* (1st ed., pp. 416–436). Cambridge University Press.

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

Bauer, B. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods, 8*(3), 338–363. https://doi.org/10.1037/1082-989X.8.3.338

Berninger, V. W., Nielson, K. H., Abbott, R. D., Wijsman, E., & Radskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology, 46*(2), 151–172. https://doi.org/10.1016/j.jsp.2007.02.007.

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

Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2013). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities, 46*(6), 490–499. https://doi.org/10.1177/0022219412468767

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., ... & DeFries, J. C. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics, 22*(3), 219–236. https://doi.org/10.1016/j.jneuroling.2008.09.003

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of*

Educational Psychology, 81(3), 313–321. https://doi.org/10.1037/0022-0663.81.3.313

Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., Cosme, M. C., & Ribeiro, I. (2017). The role of word recognition, oral reading fluency and listening comprehension in the simple view of reading: A study in an intermediate depth orthography. *Reading and Writing, 30*(3), 591–611. https://doi.org/10.1007/s11145-016-9691-3

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading, 23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

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

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*(2), 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*(1), 166–181. https://doi.org/10.1037/a0025323

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*(2), 151–164. https://doi.org/10.1177/002221940303600208

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212. https://doi.org/10.1007/BF01246098

Chall, J. S., & Jacobs, V. A. (2003). Poor children's fourth grade slump. *American Educator, 27*, 14–15.

Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading, 10*(4), 331–362. https://doi.org/10.1207/s1532799xssr1004_1

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., … Olson, R. K. (2015). Genetic and environmental etiologies of the longitudinal relations between pre-reading skills and reading. *Child Development, 86*(2), 342– 361. https://doi.org/10.1111/cdev.12295

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing, 26*(7), 1059–1086. https://doi.org/10.1007/s11145-012-9406-3

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2019). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and RAN, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading*. https://doi.org/10.1080/10888438.2019.1622546

Clinton, V., Seipel, B., van den Broek, P., McMaster, K. L., Kendeou, P., Carlson, S., & Rapp, D. N. (2014). Gender differences in inference generation by fourth-grade students. *Journal of Research in Reading, 37*(4), 356–374. https://doi.org/10.1111/j.1467‐9817.2012.01531.x

Compton, D. L. (2003). Modeling the relationship between growth in rapid naming speed and growth in decoding skill in first-grade children. *Journal of Educational Psychology, 95*(2), 225–239. https://doi.org/10.1037/0022-0663.95.2.225

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311

de Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading, 15*(4), 314–337. https://doi.org/10.1080/10888438.2010.485624

de Zeeuw, E. L., van Beijsterveldt, C. E., Glasner, T. J., de Geus, E. J., & Boomsma, D. I. (2016). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. *Learning and Individual Differences, 47*, 156–166. https://doi.org/10.1016/j.lindif.2016.01.009

Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471–479.

Dietrich, S. E., Assel, M. A., Swank, P., Smith, K. E., & Landry, S. H. (2006). The impact of early maternal verbal scaffolding and child language abilities on later decoding and reading comprehension skills. *Journal of School Psychology, 43*(6), 481–494. https://doi.org/10.1016/j.jsp.2005.10.003.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry 56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Eccles, J. S. (2005). Influences of parents' education on their children's educational attainments: The role of parent and child perceptions. *London Review of Education, 3*(3), 191–204. https://doi.org/10.1080/14748460500372309

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262–300. https://doi.org/10.3102/0034654308325998

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2017). Orthographic mapping and literacy development revisited. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 169–190). John Benjamins. https://doi.org/10.1075/swll.15.08ehr

Eklund, K., Torppa, M., & Lyytinen, H. (2013). Predicting reading disability: Early cognitive risk and protective factors. *Dyslexia, 19*(1), 1–10. https://doi.org/10.1002/dys.1447.

Eklund, K., Torppa, M., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Early cognitive predictors of PISA reading in children with and without family risk for dyslexia. *Learning and Individual Differences, 64*, 94–103. https://doi.org/10.1016/j.lindif.2018.04.012

Esmaeeli, Z., Kyle, F. E., & Lundetræ, K. (2019). Contribution of family risk, emergent literacy and environmental protective factors in children's reading difficulties at the end of second-grade. *Reading and Writing*, 1–25. https://doi.org/10.1007/s11145-019-09948-5

Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of late-emerging reading disabilities. *Journal of Learning Disabilities, 49*(1), 21–35. https://doi.org/10.1177/0022219414522706

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K. E., Michaels, R., & Shaywitz, S. E. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *The Journal of Pediatrics, 167*(5), 1121–1125.e2. https://doi.org/10.1016/j.jpeds.2015.07.045.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies?. *Educational Psychology Review, 23*(4), 553–576. https://doi.org/10.1007/s10648-011-9175-6

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*(2), 98–108. https://doi.org/10.1177/00222194050380020101

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. https://doi.org/10.1207/s1532799xssr0201_2

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*(3), 566–580. https://doi.org/10.1037/0022-0663.100.3.566

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2016). The anatomy of the RAN-reading relationship. *Reading and Writing. 29*(9), 1793–1815. https://doi.org/10.1007/s11145-016-9653-9

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing: An Interdisciplinary Journal, 25*(2), 321–346. https://doi.org/10.1007/s11145-010-9271-x

Gillett, N., Vallerand, R. J., & Lafrenière, M. K. (2011). Intrinsic and extrinsic school motivation as a function of age: The mediating role of autonomy support. *Social Psychology of Education, 15*(1), 77–95. https://doi.org/10.1007/s11218-011-9170-2.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Erlbaum

Gough, P. B., & Tumner, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Greven, C. U., Harlaar, N., Dale, P. S., & Plomin, R. (2011). Genetic overlap between ADHD symptoms and reading is largely driven by inattentiveness rather than hyperactivity-impulsivity. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 20*(1), 6-14.

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*(4), 431–447. https://doi.org/10.1353/dem.2000.0005

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102. https://doi.org/10.1037/a0038889

Hart, S. A., Petrill, S. A., DeThorne, L. S., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., & Cutting, L. E. (2009). Environmental influences on the longitudinal covariance of expressive vocabulary: Measuring the home literacy environment in a genetically sensitive design. *Journal of Child Psychology and Psychiatry, 50*(8), 911–919. https://doi.org/10.1111/j.1469-7610.2009.02074.x

Hart, S. A., Petrill, S. A., & Thompson, L. A. (2010). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences, 20*(2), 63–69. https://doi.org/10.1016/j.lindif.2009.10.004

Hawke, J. L., Olson, R. K., Willcutt, E. G., Wadsworth, S. J., & DeFries, J. C. (2009). Gender ratios for reading difficulties. *Dyslexia, 15*(3), 239–242. https://doi.org/10.1002/dys.389

Hawke, J. L., Wadsworth, S. J., & DeFries, J. C. (2006). Genetic influences on reading difficulties in boys and girls: The Colorado Twin Study. *Dyslexia, 12*(1), 21–29. https://doi.org/10.1002/dys.301

Hawke, J. L., Wadsworth, S. J., Olson, R. K., & DeFries, J. C. (2007). Etiology of reading difficulties as a function of gender and severity. *Reading and Writing*, *20*(1-2), 13–25. https://doi.org/10.1007/s11145-006-9016-z

Hjetland, H. N., Lervåg, A., Lyster, S. A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to reading comprehension: A longitudinal study from 4 to 9 years of age. *Journal of Educational Psychology, 111*(5), 751–763. https://doi.org/10.1037/edu0000321

Hogan, T., Bridges, M. S., Justice, L. M., & Cain, K. (2011). Increasing higher level language skills to improve reading comprehension. *Focus on Exceptional Children, 44*(3), 1–20.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science, 26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7*(1), 1–5. https://doi.org/10.1111/cdep.12005

Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: developmental disorders. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 369*(1634), 20120395. https://doi.org/10.1098/rstb.2012.0395

Jimenez, J. E., de Garcia, L. C., Siegel, L. S., O'Shanahan, I., Garcia, E., & Rodriguez, C. (2011). Gender ratio and cognitive profiles in dyslexia: A cross-national study. *Reading and Writing, 24*(7), 729–747. https://doi.org/10.1007/s11145 - 009 - 9222 - 6.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4*, 103–147. https://doi.org/10.1017/S0142716400004380

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Special Issue: Reading and Genetics, 29*(1), 75–91. https://doi.org/10.1111/j.1467-9817.2006.00293.x

Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research and Practice, 29*(1), 10–16. https://doi.org/10.1111/ldrp.12025

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765–778. https://doi.org/10.1037/a0015956

Kershaw, S., & Schatschneider, C. (2012). A latent variable approach to the simple view of reading. *Reading and Writing, 25*(2), 433–464 https://doi.org/10.1007/s11145-010-9278-3

Kikas, E., Pakarinen, E., Soodla, P., Peets, K., & Lerkkanen, M. K. (2018). Associations between reading skills, interest in reading, and teaching

practices in first grade. *Scandinavian Journal of Educational Research, 62*(6), 832–849. https://doi.org/10.1080/00313831.2017.1307272

Kim, Y.-S. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly, 50*(4), 459–481. https://doi.org/10.1002/rrq.107

Kim, Y.-S., Wagner, R. K., & Foster, E. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first grade readers. *Scientific Studies of Reading, 15*(4), 338–362. https://doi.org/10.1080/10888438.2010.493964

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review, 95*, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45*(3), 341–362. https://doi.org/10.1598/RRQ.45.3.4

Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poskiparta, E., Ahonen, T., Poikkeus, A.-M., & Nurmi, J.-E. (2013). The role of reading disability risk and environmental protective factors in students' reading fluency in grade 4. *Reading Research Quarterly, 48*(4), 349–368. https://doi.org/10.1002/rrq.53

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162–175. https://doi.org/10.1037/a0029285

Koponen, T., Salmi, P., Torppa, M., Eklund, K., Aro, T., Aro, M., ... & Nurmi, J.-E. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemporary Educational Psychology, 44,* 83–94. https://doi.org/10.1016/j.cedpsych.2016.02.004

Korat, O. (2009). The effect of maternal teaching talk on children's emergent literacy as a function of type of activity and maternal education level. *Journal of Applied Developmental Psychology, 30*(1), 34–42. https://doi.org/10.1016/j.appdev.2008.10.001

Korat, O., Klein, P., & Segal-Drori, O. (2007). Maternal mediation in book reading, home literacy environment, and children's emergent literacy: A comparison between two social groups. *Reading and Writing, 20*(4), 361–398. https://doi.org/10.1007/s11145/006-9034-x

Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P. F., Desrochers, A., Manolitsis, G., ... & Georgiou, G. K. (2019). Phonological awareness and rapid automatized naming as longitudinal predictors of reading in five alphabetic orthographies with varying degrees of consistency. *Scientific Studies of Reading, 23*(3), 220–234. https://doi.org/10.1080/10888438.2018.1510936

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

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Language and Reading Research Consortium. (2017). Oral language and listening comprehension: Same or different constructs*?. Journal of Speech, Language, and Hearing Research, 60*(5), 1273–1284. https://doi.org/10.1044/2017_JSLHR-L-16-0039

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211–224. https://doi.org/10.1037/0022-0663.95.2.211

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, P. (2005). Development of and relationship between phonological and motivational processes and naming speed in predicting word recognition in grade 1. *Scientific Studies of Reading, 9*(4), 367–399. https://doi.org/10.1207/s1532799xssr0904_3

Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., Siekkinen, M., Niemi, P., & Nurmi, J.-E. (2010). Luku- ja kirjoitustaidon sekä motivaation kehitys esi- ja alkuopetusvuosina [The development of reading and spelling skills from kindergarten to Grade 2]. *Kasvatus*, *41*, 116–128.

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004). Predicting reading performance during the first and the second year of primary school. *British Educational Research Journal, 30*(1), 67–92. https://doi.org/10.1080/01411920310001629974

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*(8), 1040–1048. https://doi.org/10.1111/j.1467-9280.2009.02405.x

Liederman, J., Kantrowitz, L., & Flannery, K. (2005). Male vulnerability to reading disability is not likely to be a myth: A call for new data. *Journal of Learning Disabilities, 38*(2), 109–129. https://doi.org/10.1177/00222194050380020201

Lietz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation, 32*(4), 317–344. https://doi.org/10.1016/j.stueduc.2006.10.002

Lipka, O., Lesaux, N., & Siegel, L. (2006). Retrospective analyses of the reading development of Grade 4 students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*(4), 364–378. https://doi.org/10.1177/00222194060390040901

Little, C. W., Hart, S. A., Quinn, J. M., Tucker - Drob, E. M., Taylor, J., & Schatschneider, C. (2017). Exploring the co-development of reading fluency and reading comprehension: A twin study. *Child Development, 88*(3), 934–945. https://doi.org/10.1111/cdev.12670

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*(3), 767–778. https://doi.org/10.1093/biomet/88.3.767

Logan, J. A. R., Hart, S. A., Cutting, L., Deater-Deckard, K., Schatschneider, C., & Petrill, S. (2013). Reading development in young children: Genetic and environmental influences. *Child Development, 84*(6), 2131–2144. https://doi.org/10.1111/cdev.12104

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596–613. https://doi.org/10.1037//OOI2-1649.36.5.596

Lukanenok, K. (2011). Relationship between rapid naming speed and reading speed as a marker of reading difficulties of Estonian children from 6 to 8 years. *Journal of Teacher Education for Sustainability, 13*(1), 113–128. https://doi.org/10.2478/v10099 - 011 - 0009 - x.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–14. https://doi.org/10.1007/s11881-003-0001-9

Lyytinen, H., Aro, M, Richardson, U., Erskine, J., Banff, A., Li, U. H., & Shu, H. (2015). Reading skills, acquisition of: Cultural, environmental, and developmental impediments. In J. D. Wright (Eds.), *International encyclopedia of the social and behavioral sciences* (2nd ed., pp. 5-11). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.23111-5

Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia, 55*(2), 166–192. https://doi.org/10.1007/s11881-005-0010-y

Manolitsis, G., Georgiou, G. K., & Parrila, R. (2011). Revisiting the home literacy model of reading development in an orthographically consistent language. *Learning and Instruction, 21*(4), 496–505. https://doi.org/10.1016/j.learninstruc.2010.06.005

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A metaanalytic review. *Psychological Bulletin, 138*(2), 322–352. https://doi.org/10.1037/a0026744

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267–296. https://doi.org/10.1037/a0021890

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education and Development, 19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Moll, K., Fussenegger, B., Willburger, E., & Landerl, K. (2009). RAN is not a measure of orthographic processing. Evidence from the asymmetric German orthography. *Scientific Studies of Reading, 13*(1), 1–25. https://doi.org/10.1080/10888430802631684

Moll, K., Kunze, S., Neuhoff, N., Bruder, J. & Schulte-Körne, G. (2014). Specific learning disorder: Prevalence and sex differences. *PloS One, 9*(7) e103537. https://doi.org/10.1371/journal.pone.0103537.

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., & Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in

five European orthographies: Is English an outlier orthography? *Learning and Instruction, 29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Mund, M., & Nestler, S. (2019). Beyond the cross-lagged panel model: Next-generation statistical tools for analyzing interdependencies across the life course. *Advances in Life Course Research, 41*, 100249. https://doi.org/10.1016/j.alcr.2018.10.002

Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*, 369–377. https://doi.org/10.1037/1082-989X.8.3.369

Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *npj Science of Learning, 2*(1), 1–4. https://doi.org/10.1038/s41539-017-0004-7

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties, 24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031–1039. https://doi.org/10.1111/j.1469-7610.2010.02254.x

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Reviews of Psychology, 63*, 427–453. https://doi.org/10.1146/annurev-psych-120710-100431

Olson, R. K., & Byrne, B. (2005). Genetic and environmental influences on reading and language ability and disability. In H. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 173–200). Lawrence Erlbaum.

Organisation for Economic Co-operation and Development (OECD). (2010). *PISA 2009 results: What students know and can do–Student performance in reading, mathematics, and science (Volume I)*. OECD Publishing.

Organisation for Economic Co-operation and Development (OECD) (2011). PISA 2009 at a glance. OECD Publishing.

Organisation for Economic Co-operation and Development (OECD) (2013). PISA 2012 results: What students know and can do – Student performance in mathematics, reading and science (Volume I, Revised edition). OECD Publishing.

Peng, P., Fuchs, D., Fuchs, L. S., Elleman, A. M., Kearns, D. M., Gilbert, J. K., ... & Patton III, S. (2019). A longitudinal analysis of the trajectories and predictors of word reading and reading comprehension development among at-risk readers. *Journal of Learning Disabilities, 52*(3), 195–208. https://doi.org/10.1177/0022219418809080

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition 101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F., & Bishop, D. V. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology, 60*, 283–306. https://doi.org/10.1146/annurev.psych.60.110707.163548

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., ... & Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of Abnormal Psychology, 121*(1), 212–224. https://doi.org/10.1037/a0025823

Perfetti, C. A. (1985). Reading ability. Oxford University Press.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Peterson, R. L., & Pennington, B. F. (2012). Developmental dyslexia. *Lancet, 379*(9830), 1997–2007. https://doi.org/10.1016/S0140-6736(12)60198-6

Petscher, Y. (2010). Meta-analysis of the relationship between attitudes towards reading and achievement in reading. *Journal of Research in Reading, 33*(4), 335–355. https://doi.org/10.1111/j.1467-9817.2009.01418

Petscher, Y., Justice, L. M., & Hogan, T. (2018). Modeling the early language trajectory of language development when the measures change and its relation to poor reading comprehension. *Child Development, 89*(6), 2136–2156. https://doi.org/10.1111/cdev.12880

Prior, M., Sanson, A., Smart, D., Oberklaid, F. (1995). Reading disability in an Australian community sample. *Australian Journal of Psychology, 47(1)*, 32–37.

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology, 116*(4), 914–929. https://doi.org/10.1016/j.jecp.2013.08.004

Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Scientific Studies of Reading, 21*(6), 498–514. https://doi.org/10.1080/10888438.2017.1346660

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H., Poikkeus, A.-M., ... & Lyytinen, H. (2008). Developmental links of very early phonological and language skills to second grade reading outcomes: Strong to accuracy but only minor to fluency. *Journal of Learning Disabilities, 41*(4), 353–370. https://doi.org/10.1177/0022219407311747

Quinn, J. M., & Wagner, R. K. (2015). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of Learning Disabilities, 48*(4), 433–445. https://doi.org/10.1177/0022219413508323

Rajchert, J. M., Zultak, T., & Smulczyk, M. (2014). Predicting reading and its improvement in the Polish national extension of the PISA study: The role of intelligence, trait and state anxiety, socio-economic status and school

type. *Learning and Individual Differences, 33,* 1–11. https://doi.org/10.1016/j.lindif.2014.04.003

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology, 29*(1-2), 104–122. https://doi.org/10.1080/02643294.2012.677420

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language and Hearing Research, 48*(2), 459–472. https://doi.org/10.1044/1092-4388(2005/031)

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language and Hearing Research, 52*(1), 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171)

Rutter, M., Caspi, A., Fergusson, D., Horwood, J. L., Goodman, R., Maughan, B., . . . & Carroll, J. (2004). Sex differences in developmental reading disability: New findings from four epidemiological studies. *Journal of the American Medical Association, 291(16)*, 2007–2012. https://doi.org/10.1001/jama.291.16.2007

Samuelsson, S., Byrne, B., Quain, P., Wadsworth, S., Corley, R., DeFries, J. C., ... & Olson, R. (2005). Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology, 97*(4), 705–722. https://doi.org/10.1037/0022-0663.97.4.705

Santos, S., Cadime, I., Viana, F. L., & Ribeiro, I. (2019). Cross-lagged relations among linguistic skills in european Portuguese: A longitudinal study. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.261

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental review*, *14*(3), 245–302. https://doi.org/10.1006/drev.1994.1010

Schatschneider, C., Wagner, R. K., Hart, S. A., & Tighe, E. L. (2016). Using simulations to investigate the longitudinal stability of alternative schemes for classifying and identifying children with reading disabilities. *Scientific Studies of Reading, 20*(1), 34–48. https://doi.org/10.1080/10888438.2015.1107072

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333–343. https://doi.org/10.1007/BF02294360

Segers, E., Damhuis, C. M., van de Sande, E., & Verhoeven, L. (2016). Role of executive functioning and home environment in early reading development. *Learning and Individual Differences, 49*, 251–259. https://doi.org/10.1016/j.lindif.2016.07.004

Senechal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 3 reading comprehension,

fluency, spelling, and reading for pleasure. *Scientific Studies of Reading, 10*(1), 59–87. https://doi.org/10.1207/s1532799xssr1001_4

Senechal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445–460. https://doi.org/10.1111/1467-8624.00417

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in european orthographies. *British Journal of Psychology, 94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., ... & Shaywitz, B. A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3*(1), 69–94. https://doi.org/10.1207/ s1532799xssr0301_4

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

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

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. *The New England Journal of Medicine, 326*, 145–150. https://doi.org/10.1056/NEJM199201163260301

Shaywitz, S. E., Shaywitz, B., Fletcher, J., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls. *Journal of the American Medical Association, 264*(8), 998–1002. https://doi.org/10.1001/jama.1990.03450080084036

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading, 18*(2), 132–138. https://doi.org/10.1111/j.1467-9817.1995.tb00079.x

Snowling, M. J. (2008). Specific disorders and broader phenotypes: The case of dyslexia. *The Quarterly Journal of Experimental Psychology, 61*(1), 142–156. https://doi.org/10.1080/17470210701508830

Snowling, M. J., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia, 18*(1), 1–15. https://doi.org/10.1002/dys.1432

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development, 74*(2), 358–373. https://doi.org/10.1111/1467-8624.7402003

Snowling, M. J., Hayiou-Thomas, E., Nash, H. M., & Hulme, C. (2019a). Dyslexia and developmental language disorder: Comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13140

Snowling, M. J., & Hulme, C. (2012). Annual Research Review: The nature and classification of reading disorders – A commentary on proposals for DSM-

5. *Journal of Child Psychology and Psychiatry, 53*(5), 593–607. https://doi.org/10.1111/j.1469-7610.2011.02495.x

Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin, 142*(5), 498–545. https://doi.org/10.1037/bul0000037

Snowling, M. J., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: A follow-up in early adolescence. *Journal of Child Psychology and Psychiatry, 48*(6), 609–618. https://doi.org/10.1111/j.1469-7610.2006.01725.x

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., & Wellcome Language and Reading Project Team. (2019b). Developmental outcomes for children at high risk of dyslexia and children with Developmental Language Disorder. *Child Development.* https://doi.org/10.1111/cdev.13216

Soden-Hensler, B., Taylor, J., & Schatschneider, C. (2012). Evidence for common etiological influences on early literacy skills in kindergarten. *Scientific Studies of Reading, 16*(5), 457–474. https://doi.org/10.1080/10888438.2011.599902

Soodla, P., Lerkkanen, M.-K., Niemi, P., Kikas, E., Silinskas, G., & Nurmi, J.-E. (2015). Does early reading instruction promote the rate of acquisition? A comparison of two transparent orthographies. *Learning and Instruction, 38*, 14–23. https://doi.org/10.1016/j.learninstruc.2015.02.002

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PloS One, 8*(3), e57988. doi: https://doi.org/10.1371/journal.pone.0057988

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development, 47*, 82–95. https://doi.org/10.1016/j.cogdev.2018.04.005

Swagerman, S. C., van Bergen, E., Dolan, C., de Geus, E. J., Koenis, M. M., Pol, H. E. H., & Boomsma, D. I. (2017). Genetic transmission of reading ability. *Brain and Language, 172*, 3–8. https://doi.org/10.1016/j.bandl.2015.07.008

Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., ... & Kere, J. (2003). A candidate gene for developmental dyslexia encodes a nuclear tetratrico-peptide repeat domain protein dynamically regulated in brain. *Proceedings of the National Academy of Sciences, 100*(20), 11553–11558. https://doi.org/10.1073/pnas.1833911100

Taylor, J., & Schatschneider, C. (2010). Genetic influence on literacy constructs in kindergarten and first grade: Evidence from a diverse twin sample. *Behavioral Genetics, 40*(5), 591–602. https://doi.org/10.1007/s10519-010-9368-7

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across

grade levels. *Journal of Research in Reading, 32*(4), 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Torppa, M., Eklund, K., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Why do boys and girls perform differently on PISA Reading in Finland? The effects of reading fluency, achievement behaviour, leisure reading and homework activity. *Journal of Research in Reading, 41*(1), 122–139. https://doi.org/10.1111/1467-9817.12103

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2011). Parental literacy predicts children's literacy: A longitudinal family risk study. *Dyslexia, 17*(4), 339–355. https://doi.org/10.1002/dys.437

Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-emerging and resolving dyslexia: A follow-up study from age 3 to 14. *Journal of Abnormal Child Psychology, 43*(7), 1389–1401. https://doi.org/10.1007/s10802-015-0003-1

Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly, 62*(2), 179–206. https://doi.org/10.13110/merrpalmquar1982.62.2.0179

Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities, 43*(4), 308–321. https://doi.org/10.1177/0022219410369096

Torppa, M., Niemi, P., Vasalampi, K., Lerkkanen, M.-K., Tolvanen, A., & Poikkeus, A.-M. (2019). Leisure reading (but not any kind) and reading comprehension support each other—A longitudinal study across grades 1 and 9. *Child Development*. https://doi.org/10.1111/cdev.13241

Torppa, M., Poikkeus, A.-M., Laakso, M., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42*(6), 1128–1142. https://doi.org/10.1037/0012 - 1649.42.6.1128

Torppa, M., Poikkeus, A.-M., Laakso, M., Tolvanen, A., Leskinen, E., Leppanen, P. H. T., . . . Lyytinen, H. (2007). Modeling the early paths of phonological awareness and factors supporting its development in children with and without familial risk of dyslexia. *Scientific Studies of Reading, 11*(2), 73–103. https://doi.org/10.1080/10888430709336554

Torppa, M., Vasalampi, K., Eklund, K., Sulkunen, S., & Niemi, P. (2020). Reading comprehension difficulty is often distinct from difficulty in reading fluency and accompanied with problems in motivation and school well-being. *Educational Psychology, 40*(1), 62–81. https://doi.org/10.1080/01443410.2019.1670334

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis.

Journal of Learning Disabilities, 45(5), 453–466. https://doi.org/10.1177/0022219411432685

Valdois, S., Bidet-Ildei, C., Lassus-Sangosse, D., Reilhac, C., N'Guyen-Morel, M. A., Guinet, E., & Orliaguet, J. P. (2011). A visual processing but no phonological disorder in a child with mixed dyslexia. *Cortex 47*(10), 1197–1218. https://doi.org/10.1016/j.cortex.2011.05.011

van Bergen, E., Bishop, D., van Zuijen, T., & de Jong, P. F. (2015). How does parental reading influence children's reading? A study of cognitive mediation. *Scientific Studies of Reading, 19*(5), 325–339. https://doi.org/10.1080/10888438.2015.1050103

van Bergen, E., de Jong, P. F., Plakas, A., Maassen, B., & van der Leij, A. (2012). Child and parental literacy levels within families with a history of dyslexia. *Journal of Child Psychology and Psychiatry, 53*(1), 28–36. https://doi.org/10.1111/j.1469-7610.2011.02418.x

van Bergen, E., de Jong, P. F., Maassen, B., & van der Leij, A. (2014a). The effect of parents' literacy skills and children's preliteracy skills on the risk of dyslexia. *Journal of Abnormal Child Psychology, 42*(7), 1187–1200. https://doi.org/10.1007/s10802-014-9858-9

van Bergen, E., de Jong, P. F., Regtvoort, A., Oort, F., van Otterloo, S., & van der Leij, A. (2011). Dutch children at family risk of dyslexia: Precursors, reading development, and parental effects. *Dyslexia, 17*(1), 2–18. https://doi.org/10.1002/dys.423

van Bergen, E., van der Leij, A., & de Jong, P. F. (2014b). The intergenerational multiple deficit model and the case of dyslexia. *Frontiers in Human Neuroscience, 8*, 346. https://doi.org/10.3389/fnhum.2014.00346

van Bergen, E., van Zuijen, T., Bishop, D., & de Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*(2), 147–160. https://doi.org/10.1002/rrq.160

van Setten, E. R. H., Hakvoort, B. E., van der Leij, A., Maurits, N. M., & Maassen, B. A. M. (2018). Predictors for grade 6 reading in children at familial risk of dyslexia. *Annals of Dyslexia, 68*(3), 181–202. https://doi.org/10.1007/s11881-018-0162-1

van Viersen, S., de Bree, E. H., & de Jong, P. F. (2019). Protective factors and compensation in resolving dyslexia. *Scientific Studies of Reading*, *23*(6), 461–477. https://doi.org/10.1080/10888438.2019.1603543

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. https://doi.org/10.1044/2016_JSLHR-L-16-0031

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades?. *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*(3), 407–423. https://doi.org/10.1002/acp.1414

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307–333. https://doi.org/10.2307/1912557

Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16*(1–2), 5–20. https://doi.org/10.1023/A:1021789804977

Wagner, R. K. (1986). Phonological processing abilities and reading. *Journal of Learning Disabilities. 19*, 623–629. https://doi.org/10.1177/002221948601901009

White, B. (2007). Are girls better readers than boys? Which boys? Which girls? *Canadian Journal of Education, 30*(2), 554–581. https://doi.org/10.2307/20466650

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 848–872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

Willcutt, E. G., Pennington, B. F., Olson, R. K., Chhabildas, N., & Hulslander, J. (2005). Neuropsychological analyses of comorbidity between reading disability and attention deficit hyperactivity disorder: In search of the common deficit. *Developmental Neuropsychology, 27*(1), 35–78. https://doi.org/10.1207/s15326942dn2701_3

Willcutt, E. G., Pennington, B. F., Olson, R. K., & DeFries, J. C. (2007). Understanding comorbidity: A twin study of reading disability and attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 144*(6), 709–714. https://doi.org/10.1002/ajmg.b.30310

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14*(1), 1–33. https://doi.org/10.1017/S0142716400010122

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302–314. https://doi.org/10.1080/00461520.2012.722805

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., … Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551–559. https://doi.org/10.1177/0956797610363406

ORIGINAL PAPERS

I

READING OUTCOMES OF CHILDREN WITH DELAYED EARLY VOCABULARY: A FOLLOW-UP FROM AGE 2 TO 16

by

Maria Psyridou, Kenneth Eklund, Anna-Maija Poikkeus & Minna Torppa, 2018

Research in Developmental Disabilities 78, 114-124

Research in Developmental Disabilities 78 (2018) 114–124

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16

Maria Psyridou[a,*], Kenneth Eklund[b], Anna-Maija Poikkeus[a], Minna Torppa[a]

[a] Department of Teacher Education, Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, Ruusupuisto, Alvar Aallon katu 9, 40014-FIN, Finland

[b] Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, Kärki, Mattilanniemi 6, 40014-FIN, Finland

ARTICLE INFO

Number of reviews completed is 2

Keywords:
Expressive vocabulary
Receptive vocabulary
Reading fluency
Reading comprehension
Reading difficulties

ABSTRACT

Background: Delays in expressive vocabulary have been associated with lower outcomes in reading.

Aim: The aim is to conduct a long-term follow-up study to investigate if early expressive vocabulary delay (late talking) predicts reading development in participants age 16 and under. We examine further if the prediction is different in the presence of family risk for dyslexia (FR) and early receptive vocabulary delay.

Methods: Expressive and receptive vocabulary skills were assessed at the age of 2–2.5 years, and reading skills in Grades 2, 3, 8 and 9 (age 8–16). The longitudinal sample consisted of 200 Finnish-speaking children, of which 108 had FR for dyslexia and 92 came from families without reading difficulties. We compared the reading development of five subgroups: 1) FR and no vocabulary delay; 2) FR and late talkers, 3) FR, late talkers and co-existing receptive vocabulary delay; 4) no FR and late talkers; and 5) no FR and no vocabulary delay.

Results: The group with FR and expressive and receptive vocabulary delay had difficulties in reading comprehension, but not in reading fluency. The late talkers without receptive vocabulary difficulties tended to become typical readers.

Conclusions and implications: Delays in early vocabulary can lead to a reading comprehension deficit, with the specification that expressive vocabulary deficit alone can alleviate in time, whereas the combined deficit is a stronger risk marker.

What this paper adds?

The present study extends the literature by investigating the reading development of children with early expressive vocabulary delay followed up to age 16 (Grade 9) in a Finnish language context in relation to both reading fluency and reading comprehension. In addition, we examine whether this relationship is different in the presence and absence of other co-occurring risk factors, in particular family risk for dyslexia and early receptive vocabulary delay. The existing literature on reading development of children with delays in early vocabulary is almost exclusively limited to findings concerning the early grades of primary school. Furthermore, all the previous studies spanning beyond Grade 2 have been conducted using English-speaking children; thus, further research in other orthographies is needed.

We compared the reading development of five subgroups: 1) FR and no vocabulary delay; 2) FR and late talkers; 3) FR, late talkers

* Corresponding author.
E-mail addresses: maria.m.psyridou@jyu.fi (M. Psyridou), kenneth.m.eklund@jyu.fi (K. Eklund), anna-maija.poikkeus@jyu.fi (A.-M. Poikkeus), minna.p.torppa@jyu.fi (M. Torppa).

https://doi.org/10.1016/j.ridd.2018.05.004
Received 14 June 2017; Received in revised form 24 February 2018; Accepted 10 May 2018
Available online 24 May 2018
0891-4222/

and co-existing receptive vocabulary delay; 4) no FR and late talkers; and 5) no FR and no vocabulary delay. Our findings add to the literature by suggesting that a trajectory of later reading comprehension difficulties is much more likely when early expressive vocabulary delay is accompanied by receptive vocabulary delay. Our findings suggest that a delay in early vocabulary can lead to a persistent deficit with the specification that expressive vocabulary deficit alone can be alleviated in time, whereas the combined deficit is a stronger risk marker. Early expressive vocabulary delay represented high risk only for reading comprehension development and only when it co-occurred with other risk factors (family risk for dyslexia and receptive vocabulary difficulties).

1. Introduction

Early years expressive language delay, or late talking, is one of the most common concerns of parents and early health care personnel, and it is one of the common reasons for referring young children for evaluation. This is not surprising, as early language development forms the foundation for later educational and academic achievement and has important links with social adaptation (Reilly et al., 2010). From a clinical and educational perspective, it is important to be able to predict as early as possible which children may be at risk for learning difficulties in their school years. Delays in expressive vocabulary have been associated with lower outcomes in reading (Rescorla, 2002, 2005, 2009), but apart from Rescorla's sample from the USA, little research has focused on literacy development after the early grades of primary school using longitudinal designs. Moreover, most of the previous studies have been conducted among English-speaking children and have focused almost solely on early expressive, but not receptive vocabulary.

The present study will extend previous studies (Duff, Reen, Plunkett, & Nation, 2015; Lyytinen, Eklund, & Lyytinen, 2005; Rescorla, 2002, 2005, 2009) to later years of schooling by examining if late talking (delays in early expressive vocabulary) predict reading development (fluency and comprehension) from early primary Grades 2 and 3 to lower secondary Grades 8 and 9 in Finland. We will also examine the effects of co-occurring receptive vocabulary delay and family risk for dyslexia on the relationship between early vocabulary skills and reading outcomes. Expressive vocabulary may not be a marker of subsequent language and literacy difficulties when it is not accompanied by co-occurring difficulties in receptive vocabulary (Lyytinen et al., 2005). It is also possible that family risk for dyslexia actually explains the link between expressive vocabulary delay because the children with family risk (due to parental language or reading difficulties) are shown to be at risk for both vocabulary delays (Duff et al., 2015; Lyytinen et al., 2005; Snowling, Gallagher, & Frith, 2003; Snowling, Muter, & Carroll, 2007) and for later language and reading outcomes (Lyytinen et al., 2005; Reilly et al., 2010).

Early vocabulary skills (expressive or a composite score of expressive and receptive vocabulary) have been shown to be associated with both reading fluency (Nation & Snowling, 2004; Snowling et al., 2003) and reading comprehension (Duff et al., 2015; Nation, Cocksey, Taylor, & Bishop, 2010; Ouellette 2006; Ouellette & Beers 2010). Theories on how early vocabulary skills and reading outcomes are linked depend, however, on the reading skill in question. According to the lexical restructuring hypothesis, the link between early vocabulary and reading fluency is indirect and mediated via phonological skills (e.g. Walley, Metsala, & Garlock, 2003). Phonological awareness plays a critical role in the development of decoding (Ehri et al., 2001; Hulme & Snowling, 2014; Melby-Lervåg, Lyster, & Hulme, 2012), and tasks requiring identification or manipulation at the phoneme level, in particular, have been found to be associated with variations in decoding skills (Georgiou, Parrila, & Papadopoulos, 2008; Lervåg, Bråten, & Hulme, 2009). Walley et al. (2003) suggested that phonological awareness may be linked to vocabulary via the process of restructuring phonological representations (the word cat, for example, first segmented at syllable level/k/-/æt/becoming segmented at phoneme level/k/-/æ/-/t/). Vocabulary growth increases sensitivity towards phonological similarities between words, which elicits a restructuring of the already-existing phonological representations in the lexicon (Walley et al., 2003).

The link between reading comprehension and vocabulary skills may be more direct than that of vocabulary and reading fluency. According to the simple view of reading model (Gough & Tumner, 1986; Hoover & Gough, 1990), reading comprehension is based on two core skills: decoding and language comprehension. Reading comprehension has been shown to rely on a variety of oral language comprehension skills (Hulme & Snowling, 2014), including expressive and receptive vocabulary (e.g. Nation et al., 2010) and listening comprehension (e.g. Catts, Adlof, & Weismer, 2006; Nation et al., 2010). Weaknesses in these skills have been found to be manifested before learning to read, thus providing a possible causal link between the influence of oral language comprehension on difficulties in reading comprehension (e.g. Catts et al., 2006; Nation et al., 2010). This association between vocabulary and reading comprehension has also been shown through intervention studies (Clarke, Snowling, Truelove, & Hulme, 2010; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013).

The vast majority of studies on the reading development of children with expressive vocabulary delay have focused on language outcomes in kindergarten (age range between 2 and 5.5 years) (Feldman et al., 2005; Moyle, Weismer, Evans, & Lindstrom, 2007; Thal, Miller, Carlson, & Vega, 2005) or reading outcomes in the early grades of primary school (Grade 2) (Lyytinen et al., 2005). The documentation on the reading development of the children with expressive vocabulary delay beyond the early grades comes from one sample (Rescorla, 2002, 2005, 2009). Rescorla (2002) showed that children that had been identified as having expressive vocabulary delay in early childhood continued to manifest significantly poorer reading skills (i.e., decoding, comprehension, written language and spelling) than their peers at the ages of 8 and 9. At the age of 13, the children with expressive vocabulary delay demonstrated significantly lower vocabulary and reading comprehension skills than the comparison children (Rescorla, 2005), whereas no differences were observed in reading fluency (Rescorla, 2005). At the age of 17, the children with expressive vocabulary delay had average reading skills but below average vocabulary and verbal memory (Rescorla, 2009). Rescorla's findings show that early expressive vocabulary delay is negatively associated with reading achievement, especially after the initial stages of learning to read (Rescorla, 2002, 2005, 2009), and the effect is primarily seen on reading comprehension rather than reading fluency (Rescorla, 2005, 2009). Rescorla, however, did not control for other risk factors or co-occurring difficulties, and it is possible that late talking is not the

cause of subsequent reading difficulties. A few studies show that expressive vocabulary delay may actually be a risk factor for reading development only when accompanied by other risk factors, such as family risk for dyslexia or language difficulties and receptive vocabulary delays (Duff et al., 2015; Lyytinen et al., 2005; Poll & Miller, 2013).

Family risk for dyslexia is linked to both vocabulary and reading difficulties and may be the underlying cause for late talkers' problems in reading (Duff et al., 2015; Lyytinen et al., 2005; Snowling et al., 2003, 2007; Torppa, Lyytinen, Erskine, Eklund, & Lyytinen, 2010). It has been shown that the combination of family risk for reading difficulties and early delays in expressive and receptive vocabulary impedes reading development in the early grades (Duff et al., 2015; Lyytinen et al., 2005). In the same Finnish sample that is used for this study, Lyytinen et al. (2005) showed that those children with expressive vocabulary delay at the end of the second grade (age 8 years) who also exhibited evidence of receptive vocabulary delay and a family risk for dyslexia were often diagnosed as manifesting reading difficulties, whereas those children with only expressive vocabulary delay were not. We add to their findings by extending the investigation to Grades 3, 8 and 9 and by examining both reading fluency and reading comprehension development from Grade 2–9.

2. The present study

The existing literature on reading development of children with delays in early vocabulary is almost exclusively limited to findings concerning the early grades of primary school, and all the studies except one (Lyytinen et al., 2005) have been conducted using English-speaking children. The present study extends the literature by investigating the reading development of children with early expressive vocabulary delay followed up to age 16 (Grade 9) with native Finnish speakers with respect to both reading fluency and reading comprehension. In addition, we examine whether this relationship is different in the presence and absence of other co-occurring risk factors: family risk for dyslexia and early receptive vocabulary delay. We use reading fluency and comprehension as outcomes rather than reading accuracy because most Finnish children can read accurately after the first year of formal education, and accuracy measures are at a ceiling by that time. The use of fluency and comprehension as indicators of reading progress is typical for languages that have a high level of transparency of the orthographic system (Aro & Wimmer, 2003; Seymour, Aro, & Erskine, 2003; Share, 2008). The Finnish orthographic system is considered very transparent (Aro, 2016), and in contrast to English, for instance, there is almost a perfect one-on-one correspondence between phonemes and graphemes, with every letter always having the same sound and every sound always being represented by the same letter. The combination of high-quality reading teaching and the high transparency of the orthography supports faster reading acquisition than in the less consistent orthographies, such as English (e.g. Seymour et al., 2003). In Finland, children start attending primary school in August of the year in which they turn 7. The follow-up period of the present study was from Grade 2 (when the children's ages varied between 8 and 9 years) to the end of Grade 9 (when their ages varied between 15 and 16 years).

The research questions of the present study were as follows: (1) Does early expressive vocabulary delay predict difficulties in reading fluency or reading comprehension at Grades 2, 3, 8 and 9? (2) Is the relationship between early expressive vocabulary delay and reading development different in the presence and absence of receptive vocabulary delay and family risk for dyslexia?

Based on previous studies (Nation & Snowling, 2004; Rose & Rouhani, 2012), we expected that vocabulary skills would be associated with reading development. However, we predicted that early vocabulary would be a better predictor for reading comprehension than reading fluency (Rescorla, 2005, 2009; Ricketts, Nation & Bishop, 2007). We also expected that children with early expressive vocabulary delay would have more reading comprehension problems in the later grades than in the earlier grades (Rescorla, 2002, 2005, 2009) and that the most severe reading problems would be found among children with both expressive and receptive vocabulary delays (Lyytinen et al., 2005). Finally, we expected that the children with family risk for dyslexia would be slower readers than the children of parents with no reading difficulties (Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015; Snowling & Melby-Lervåg, 2016).

3. Method

3.1. Participants

The children (n = 200) were participants of the "Jyväskylä Longitudinal Study of Dyslexia (JLD)", a longitudinal family risk study following children with and without family risk for dyslexia. The participants were selected from among families of 9368 newborns born in the province of "central Finland" between April 1993 and July 1996. The selection of the parents with dyslexia followed a three-step procedure. The first stage was a short questionnaire including three questions concerning difficulties in learning to read and spell among themselves and their close relatives. The second stage was a more detailed questionnaire, which focused on demographic information and the occurrence of reading and writing difficulties during childhood, adolescence and among relatives. Those who fulfilled the criteria of inclusion in the first and second stage were invited for an interview and for an assessment of their reading and writing skills to confirm their present status of dyslexia. The criteria used to select the parents with dyslexia were as follows: a self-reported childhood history of reading or writing difficulties, self-reported adulthood situation of reading and writing difficulties, and present performance in diagnostic tasks of reading and writing (for full details of recruitment, see Leinonen et al., 2001). The criteria for a child to be included in the family risk group (FR, n = 108) included that one of the parents was identified as having reading difficulties in the reading and spelling tests, literacy problems during early school years, and at least one first-degree relative with corresponding difficulties. In the control group without family risk, neither parent had reported reading difficulties or a family history of dyslexia, and neither had difficulties in reading and spelling tasks (NR, n = 92). The parents' educational

backgrounds were assessed using a 7-point scale: 1 = comprehensive school without any vocational education, 2 = comprehensive school with short-term vocational courses, 3 = comprehensive school with a vocational school degree, 4 = comprehensive school with a vocational college degree, 5 = comprehensive school with a lower university degree (Bachelor's) or a degree at a polytechnic, 6 = upper secondary school with a BA degree or a degree at a polytechnic, and 7 = upper secondary school with a higher university degree (master's or a doctorate degree). The mean of the mothers' education was 4.18 (SD = 1.48) in the FR group and 4.52 in the control group (SD = 1.35). The mean of the fathers' education was 3.68 (SD = 1.27) and 3.80 (SD = 1.40), respectively. There were no differences in parental education or in their nonverbal IQs. The IQ of all parents, assessed with the Raven B, C and D matrices, was ≥80 (Raven, Court, & Raven, 1992).

3.2. Measures

The children were assessed at ages 24 months, 30 months, 8–9 years (Grade 2), 9–10 years (Grade 3), 14–15 years (Grade 8) and 15–16 years (Grade 9). Trained testers individually assessed the children's language skills at ages 24 and 30 months in a laboratory setting, and school-age reading fluency and reading comprehension were assessed via group-administered tests in the classrooms. Reading fluency was assessed with three tasks: oral text reading, oral pseudoword text reading and oral word list reading. Reading comprehension was assessed using a short passage reading comprehension task and PISA reading comprehension (Grade 9). The measures are described in detail below.

3.2.1. Expressive vocabulary measures at the age of 24 and 30 months

The composite score of expressive vocabulary was calculated based on the following four measures: vocabulary production and maximum sentence length using the Finnish adaptation of the MacArthur Communicative Development Inventory (Fenson et al., 1994) at the age of 24 months using parental reports, the Bayley Scales of Infant Development (BSID; Bayley, 1993) expressive score (naming pictures and objects) at the age of 24 months and the expressive score of the Reynell Developmental Language Scale (RDLS; Reynell & Huntley, 1987) at the age of 30 months. A z-score based on control group distribution was calculated for each measure, and the average of the z-scored measures served as the composite score for expressive language. The Cronbach alpha reliability for the composite score of expressive language was 0.86.

3.2.2. Receptive vocabulary

The children's verbal comprehension level was assessed using the Reynell Receptive Language Scale at the age of 30 months.

3.2.3. Oral text reading (Grades 2, 3 and 8)

In the spring term of each grade level, oral text reading was assessed by asking participants to read an age-appropriate text aloud. In Grade 2, the children read a text ("Exciting journeys") consisting of 19 sentences, with a total of 124 words; in Grade 3, they read a text ("Useless belongings") consisting of 18 sentences, with a total of 189 words; and in Grade 8, they read a text ("Fields of Lapland") consisting of 16 sentences, with a total of 207 words. Children's reading performances were recorded, and the accuracy of reading and time spent on reading were checked. The scores of the children's speed were measured as words per minute.

3.2.4. Oral pseudoword text reading (Grades 2, 3 and 8)

The participants read aloud a short text that was made up of 19 pseudowords (Grade 2) or 38 pseudowords (Grades 3 and 8). The pseudowords and sentence structures resembled real Finnish in form, but had no meaning. The mean word length was 7.21 letters/word in Grade 2 and 7.29 letters/word in Grades 3 and 8. Children's reading performances were recorded, and the accuracy of reading and time spent on reading were checked. The scores of the children's speed were measured as pseudowords per minute.

3.2.5. Oral word list reading (Grades 2, 3 and 8)

In the nationally normed Lukilasse reading test (Häyrinen, Serenius-Sirve, & Korkman, 1999), the participants had 2 min to read aloud as many words as possible from a 90-item (Grade 2) or 105-item (Grade 3) list of words assembled vertically in columns. The same list which was used in Grade 3 was administered in Grade 8, but the time limit was reduced to 1 min. The length of the words increased gradually, ranging from 3 to 18 letters/word in Grade 2 and from 3 to 22 letters/word in Grades 3 and 8. The mean length of the words was 9.08 letters in Grade 2 and 9.57 letters in Grades 3 and 8. The children received 1 point for each correct answer.

3.2.6. Reading comprehension (Grades 2 and 3)

A group-administered subtest of the nationally normed reading test battery (Ala-asteen lukutesti, or ALLU; Lindeman, 2000) was used. The children silently read a fiction story and then answered 12 questions (11 multiple choice questions and one that required arranging five statements into the correct sequence). The length of the text in Grade 2 was five paragraphs and had 114 words. In Grade 3, the length of the text was 4 paragraphs and had 139 words. The children received 1 point for each correct answer (max = 12). The maximum time allotted was 45 min. The Kuder–Richardson reliability coefficients reported by Lindeman (2000) were 0.80 in Grade 2 and 0.75 in Grade 3.

3.2.7. Reading comprehension (Grade 9): PISA reading

The reading tasks consisted of the Programme for International Student Assessment (PISA) reading link items, which are contained in each cycle of the PISA assessments to provide comparability of the measurement. It is a triennial international survey which

aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students (OECD, 2010, p. 26; OECD, 2013, p. 45). The students were given a booklet with eight texts to read, and they had to answer questions in the allotted space. The reading materials included texts, tables, graphs and figures. The total length of the material is 30 A4 pages. However, because the material includes many pictures and tables, the estimated length of the text is approximately 10 pages. There were 15 multiple-choice questions and 16 questions that required written responses. Of the questions, 12 required students to access and retrieve information, 12 to integrate and interpret information, and 7 to reflect and evaluate information. Students had 60 min to complete the task. A total score was calculated representing all the PISA reading items. The Cronbach's alpha reliability coefficient for the total score in this sample was 0.80.

3.3. Statistical analysis

First, all the variable distributions were examined. In grade 8, two extreme outliers were observed in the pseudoword text reading and three in the text reading and skewness values were rather high. Outliers were moved to the tails of the distributions, and log10 transformations were performed. After the transformations, the distributions approximated normal distribution. In the examination of reading fluency, a composite variable for each grade was calculated by converting the scores for each reading fluency measure into z-scores and by calculating mean composite scores. Cronbach alpha reliability for the fluency composite was 0.91, 0.88 and 0.90 in Grades 2, 3 and 8, respectively.

One-way ANOVAs were conducted to compare the five groups in reading fluency and reading comprehension performance. The groups were identified with the following criteria: Expressive vocabulary delay was identified based on three tests at 2–2.5 years: 1) vocabulary production and maximum sentence length from the Finnish adaptation of the CDI (MacArthur Communicative Development Inventory) 2) the Bayley expressive score (Bayley, 1993), and 3) the expressive score from the Reynell Developmental Language Scales (Reynell & Huntley, 1987). The Reynell Receptive Language Scale scores from the RDLS were used to further divide late talkers (i.e., children identified with expressive vocabulary delay) into groups with and without receptive vocabulary delay. Of the five subgroups, three were with family risk and two with no family risk: 1) family risk and no vocabulary delay (FR); 2) family risk and late talkers (FR-LT1); 3) family risk, late talkers and receptive vocabulary delay (FR-LT2); 4) no family risk and late talkers (NR-LT1), and 5) no family risk and no vocabulary delay (NR).

There were two groups of late talkers (FR-LT1 and NR-LT1) who had only expressive vocabulary delay. They had expressive vocabulary score at least one standard deviation below the mean of children in the control group but age-appropriate receptive skills. There were 10 children with family risk for dyslexia (FR-LT1) and 10 children with no risk for dyslexia (NR-LT1). The second subtype of late talkers demonstrated both an expressive and receptive vocabulary delay (at least one standard deviation below the mean on both scores). There were 12 children from the family risk sample (FR-LT2), but only three children from the no-risk sample fulfilled these criteria, and they were subsequently excluded from the analyses. Finally, the analyses involved two groups with no vocabulary delay: the FR group (N = 83) and the NR group (N = 79) with age-appropriate or better early expressive and receptive vocabulary. In addition to the group comparisons, we also examined the correlations between all measures.

The descriptive statistics of the measures used in the identification of the five groups are presented in Table 1. Three children of the family risk sample were not included in the analysis because the Reynell expressive and receptive scores were missing for two of them and because one of the children attended school one year later than his age-mates and we did not have his reading measure outcomes. None of the children had reports on Specific Language Impairment (SLI) diagnoses at any age. In Finland, physicians use the International Classification of Diseases (ICD) when diagnosing specific language impairment (SLI) (Asikainen, 2005). However, some milder language or language-related difficulties were diagnosed in 11 children (one had naming difficulty, five demonstrated delayed language/speech development, four had dysphasic features, and one faced learning difficulties, especially in language). Of these 11 children, six belonged to the group with family risk and both expressive and receptive vocabulary delays (FR-LT2); three belonged to the group with family risk, but no early vocabulary delays (FR); and two belonged to the control group with expressive vocabulary delay (NR-LT1).

4. Results

4.1. Group differences in reading fluency

The one-way ANOVAs for reading fluency showed significant group differences in Grades 2, 3 and 8 (Tables 2 and 3, Fig. 1). Pairwise comparisons (LSD) revealed that in Grades 2 and 3 the NR group children were faster readers than the FR group children. In grade 3, the NR group children were also faster readers than the FR-LT2 group children. In Grade 8, the pairwise comparisons (Dunnett T3) showed that the NR group children were faster readers than the FR group children. Further inspection of the effect sizes showed that the comparisons between the three groups with family risk for dyslexia and the two groups without family risk for dyslexia all had small to medium effect sizes.

4.2. Group differences in reading comprehension

The one-way ANOVAs for reading comprehension showed significant group differences in Grades 2 and 9, but not in Grade 3 (Tables 2 and 3, Fig. 2). In Grade 2, the pairwise comparisons (Dunnett T3) showed no significant differences between the groups but the effect sizes were large in favour of the NR group compared to the FR-LT2 and NR-LT1. In Grade 3, the pairwise comparisons (LSD)

Table 1
Descriptive statistics for the vocabulary measures used to determine the subgroups.

	Family risk, no language delay (FR) (N = 83)		Family risk, Late talkers (FR-LT1) (N = 10)		Family risk, Late talkers and receptive language delay (FR-LT2) (N = 12)		No family risk, Late talkers (NR-LT1) (N = 10)		No family risk, no language delay (NR) (N = 79)			
	M	SD	M	SD	M	SD	M	SD	M	SD	F	df
2 years												
Vocabulary production (CDI)	305.92	135.79	57.70	39.81	87.25	72.88	33.10	39.24	321.36	138.50		
Maximum sentence length (CDI)	5.67	2.24	2.10	1.12	2.70	1.47	1.63	1.30	6.47	2.67		
Bayley expressive vocabulary	10.32	3.51	1.40	1.95	3.18	3.52	1.10	2.33	10.97	2.89		
2.5 years												
Reynell Receptive Language Scale	37.07^{2}	6.07	36.50^{2}	2.27	26.75^{1}	3.86	36.40^{2}	2.99	37.91^{2}	5.83	10.40	4,184***
Reynell Developmental Language Scale	36.05	5.00	28.50	2.84	24.08	5.23	22.60	8.47	35.87	4.55		
Composite score for expressive vocabulary	0.15^{2}	0.57	-1.31^{1}	0.17	-1.27^{1}	0.49	-1.63^{1}	0.46	0.28^{2}	0.58	56.44	4,189***

Notes: Groups with different subscripts (^{1,2}) differed from each other in post-hoc comparisons, using LSD. The composite score for expressive vocabulary was calculated as average of the standardized scores for the a) Vocabulary production, b) Maximum sentence length, c) Bayley expressive vocabulary and d) Reynell Developmental Language Scale. Standardized scores are based on the distribution of the control group. The five groups (FR, FR-LT1, FR-LT2, NR-LT1, NR) were identified with respect to status of expressive and receptive vocabulary delay and family risk for dyslexia. The FR-LT1 (N = 10) and the NR-LT1 (N = 10) groups had at least 1 SD below the mean of children in the control group in the composite score of expressive vocabulary but age-appropriate receptive skills. The FR-LT2 (N = 12) group demonstrated both an expressive and receptive vocabulary delay (at 1 SD below the mean on both scores). The FR (N = 83) and the NR (N = 79) groups had age-appropriate or better early expressive and receptive vocabulary. *$p \leq 0.05$, **$p \leq 0.01$. ***$p \leq 0.001$.

Table 2
Descriptive statistics and group comparisons for reading fluency and reading comprehension.

	FR		FR-LT1		FR-LT2		NR-LT1		NR			
	M	SD	M	SD	M	SD	M	SD	M	SD	F	df
Grade 2												
Reading Fluency ^{z}	-0.26^{1}	0.95	$-0.18^{1,2}$	1.05	$-0.19^{1,2}$	0.98	$0.21^{1,2}$	1.01	0.36^{2}	0.95	4.40**	4,185
Reading comprehension	8.99	3.00	8.44	2.92	6.55	3.33	7.00	2.55	9.77	1.79	5.27***	4,166
Grade 3												
Reading fluency ^{z}	-0.22^{1}	0.98	$-0.28^{1,2}$	1.25	-0.32^{1}	0.93	$0.05^{1,2}$	1.30	0.33^{2}	0.90	3.72**	4,189
Reading comprehension	9.76^{1}	1.76	10.20^{1}	1.32	8.64^{2}	2.01	$10.00^{1,2}$	2.78	10.15^{1}	1.40	2.12	4,170
Grade 8												
Reading fluency ^{z}	-0.23^{1}	1.13	$-0.28^{1,2}$	0.96	$-0.22^{1,2}$	0.85	$-0.03^{1,2}$	1.55	0.37^{2}	0.68	3.95**	4,174
Grade 9												
Reading comprehension	0.12^{1}	0.79	-0.17^{1}	0.73	-1.14^{2}	1.12	-0.15^{1}	1.09	0.25^{1}	0.90	5.91***	4,153

Notes: Groups with different subscripts (^{1,2}) differed from each other in post-hoc comparisons, using either LSD or Dunnett T3, depending on equality of the variances. Subscript z refers to standardized score. The reading fluency measures are composite scores calculated as average of the standardized scores for the three reading fluency tasks. Standardized scores are based on the distribution of the control group.*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

Table 3
Effect sizes for group comparisons in reading fluency and reading comprehension.

	Effect size[a]									
	FR vs FR-LT1	FR vs FR-LT2	FR vs NR-LT1	FR vs NR	FR-LT1 vs FR-LT2	FR-LT1 vs NR-LT1	FR-LT1 vs NR	FR-LT2 vs NR-LT1	FR-LT2 vs NR	NR-LT1 vs NR
Grade 2										
Reading fluency	0.08	0.07	0.50	0.65	0.01	0.38	0.54	0.40	0.57	0.15
Reading comprehension	0.19	0.77	0.72	0.31	0.60	0.53	0.55	0.15	**1.20**	**1.26**
Grade 3										
Reading fluency	0.05	0.10	0.23	0.58	0.04	0.26	0.56	0.33	0.71	0.25
Reading comprehension	0.29	0.60	0.10	0.25	**0.92**	0.09	0.04	0.56	**0.88**	0.07
Grade 8										
Reading fluency	0.05	0.01	0.15	0.64	0.07	0.19	0.78	0.15	0.77	0.33
Grade 9										
Reading comprehension	0.37	**1.30**	0.30	0.16	**1.03**	0.02	0.51	**0.90**	**1.37**	0.40

Note: Large (> 0.80) effect size with bold.
[a] The value of Cohen's d was calculated using the means and standard deviations of two groups. Cohen d is considered large when $d > 0.80$, medium when $d > 0.50$ and small when $d > 0.30$.

suggested that children in the FR-LT2 group were poorer comprehenders than children in the FR, FR-LT1 and NR groups. The effect sizes were medium for the FR-LT2 comparison with FR and large in comparison with the FR-LT1 and NR groups. In Grade 9, the pairwise comparisons (LSD) showed lower reading comprehension scores for the FR-LT2 group than for all the other groups, and these differences were accompanied by large effect sizes.

4.3. Correlations between early expressive and receptive vocabulary and literacy measures

Correlations (Table 4) showed that expressive and receptive vocabulary measures were strongly associated. Both expressive and receptive vocabulary had stronger correlations with reading comprehension than reading fluency. The correlations were non-significant between the vocabulary measures and Grade 8 reading fluency. In contrast, the correlations between expressive and receptive vocabulary and reading comprehension were statistically significant across all grades.

5. Discussion

The reading outcomes of children with expressive vocabulary delay were investigated in this study by analyzing their reading development up to 16 years of age using measures of both reading fluency and comprehension. The role of the additional risk factors of family risk for dyslexia and poor early receptive vocabulary were also examined. Five subgroups were identified based on children's early expressive vocabulary, receptive vocabulary and family risk status, and they were compared in reading fluency and comprehension measures in Grades 2, 3, 8 and 9. In general, the findings suggest that early expressive language delay predicts reading comprehension, but not reading fluency development. Late talking alone was not, however, a sufficient risk index for reading comprehension difficulties for children with family risk for dyslexia or for children without such risk. Only the children with both expressive and receptive vocabulary delays and family risk had clear difficulties in reading comprehension, and the difficulties were

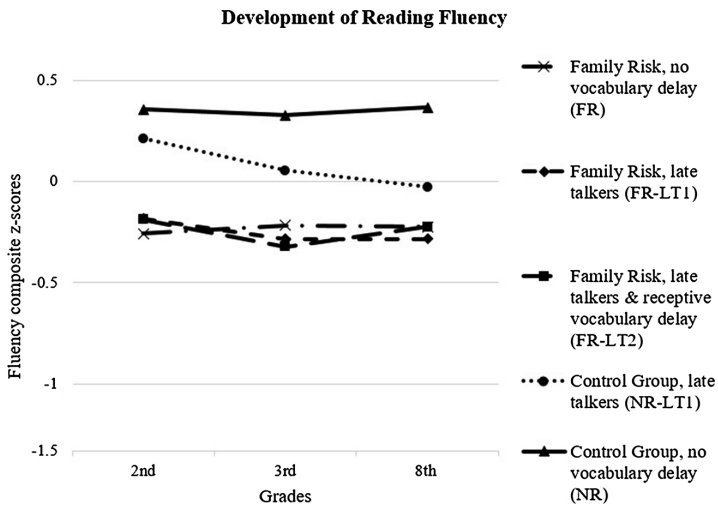

Fig. 1. Development of fluency skills by group for each grade.

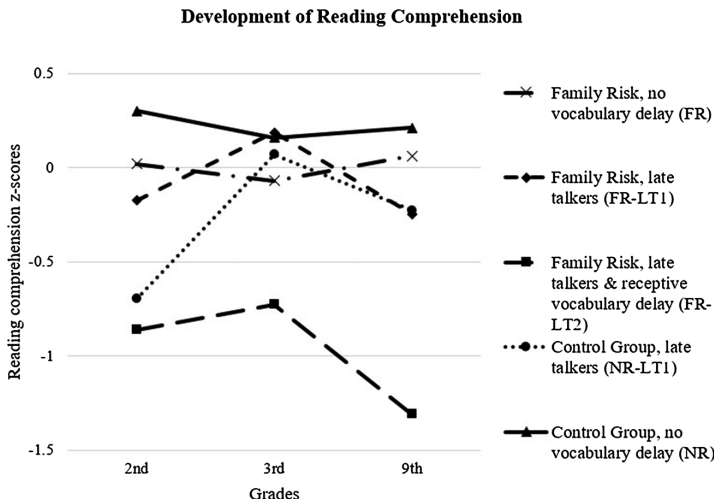

Fig. 2. Development of reading comprehension by group for each grade.

sustained to adolescence.

The children who had delays in both expressive and receptive vocabulary as toddlers demonstrated persistent weaknesses in comparison to the other groups in reading comprehension measures. As indicated by moderate to large effect sizes, these children had lower scores compared to all the other groups at the end of Grade 2 and 3, and they continued on a declining trajectory to the end of Grade 9. Although the comparison between all the groups in Grade 3 did not reveal statistically significant differences, the differences were clear in pairwise comparisons and in the effect sizes of the group comparisons, which become larger from Grade 3 onwards. All the groups except FR-LT2 performed at a similar level in Grade 3.

Our findings add to the literature (Duff et al., 2015; Lyytinen et al., 2005; Rescorla, 2002, 2005, 2009) by suggesting that a trajectory of later reading comprehension difficulties is much more likely if early expressive vocabulary delays are accompanied by receptive vocabulary delays. In our sample, the children with both receptive and expressive vocabulary delay also had family risk for dyslexia, which makes it difficult to separate the effects of family risk and receptive vocabulary delay. However, because the other family risk children (the group who did not have a vocabulary delay and the group who had only an expressive delay) manifested average reading comprehension, it can be assumed that it is the combined vocabulary difficulty (expressive and receptive), rather

Table 4
Correlation table for expressive and receptive vocabulary and reading fluency and reading comprehension in grades 2, 3, 8, and 9.

		1.	2.	3.	4.	5.	6.	7.
1.	Expressive vocabulary							
2.	Receptive vocabulary	0.49***						
3.	Reading fluency, Grade 2	0.16*	0.18*					
4.	Reading fluency, Grade 3	0.20**	0.23**	0.89***				
5.	Reading Fluency, Grade 8	0.11	0.12	0.71***	0.79***			
6.	Reading comprehension, Grade 2	0.39***	0.34***	0.48***	0.44***	0.40***		
7.	Reading comprehension, Grade 3	0.19*	0.26***	0.35***	0.36***	0.38***	0.45***	
8.	Reading comprehension, grade 9	0.32***	0.30***	0.29***	0.40***	0.33***	0.37***	0.45***

* $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

than family risk per se, that underlies reading comprehension difficulties. Moreover, the fact that the combination of receptive and expressive vocabulary delays was more common in the family risk group (12 children) than in the no-risk group (only three children) suggests a link between family risk for reading difficulties and combined vocabulary delay.

The findings suggest that a delay in early vocabulary can lead to a reading comprehension deficit, which is in line with suggestions in the previous studies (Duff et al., 2015; Lyytinen et al., 2005; Rescorla, 2005, 2009), with the specification that an expressive vocabulary deficit alone can be alleviated in time, whereas the combined deficit is a stronger risk marker. Previous studies have also shown correlative evidence for the predictive association between language skills and reading comprehension (Catts et al., 2006; Duff et al., 2015; Nation et al., 2010). The present findings are in line with Rescorla's (2005) findings indicating that children with delays in early vocabulary demonstrated no differences in basic reading mechanisms (including fluency) at age 13, but they had significantly lower scores in reading comprehension. In addition, they are in line with the findings of Duff et al. (2015) suggesting that early vocabulary (in their study combination of expressive and receptive) has a stronger relationship with reading comprehension than reading accuracy. Because early vocabulary was measured at the age of 2–2.5 years in the present study, which was long before reading had started to develop and which prevented any concerns about the direction of effects, a predictive association between early vocabulary and reading comprehension outcomes could be suggested. This link between reading comprehension and early vocabulary could be explained by the simple view of reading model (Gough & Tumner, 1986; Hoover & Gough, 1990). Reading comprehension has been shown to rely on a variety of oral language comprehension skills (Hulme & Snowling, 2014), including expressive and receptive vocabulary (e.g. Nation et al., 2010) and listening comprehension (e.g. Catts et al., 2006; Nation et al., 2010). Weaknesses in these skills have been found to be manifested before learning to read, which could provide a possible causal link between the influence of oral language comprehension on difficulties in reading comprehension (e.g. Catts et al., 2006; Nation et al., 2010).

In contrast to reading comprehension, reading fluency was found to be linked with family risk for dyslexia and not with the presence of early expressive vocabulary delay. The finding that family risk was predictive of reading fluency was expected and supports previous studies indicating the strong effects of family risk in developmental dyslexia (Lyytinen et al., 2015; Snowling et al., 2003; van Bergen, de Jong, Plakas, Maassen, & van der Leij, 2012). Based on our findings, it is thus likely that children with reading difficulties who show early vocabulary difficulties have additional risk factors, such as family risk (Nation & Snowling, 2004; Snowling et al., 2003; Torppa et al., 2010).

The finding of no link from late talking to reading fluency development appears to contradict the lexical restructuring hypothesis (e.g. Walley et al., 2003), which suggests a link between early vocabulary and reading fluency mediated via phonological skills. In addition, we have shown earlier in this same sample that early vocabulary is linked via phonological skills to kindergarten precocious reading ability (Torppa et al., 2007), supporting the lexical restructuring hypothesis. However, in the present study, we focused on a phase in reading development where phonological skills may not be strong predictors of reading fluency anymore, at least in the context of a transparent orthography (see e.g. Aarnoutse, van Leeuwe, & Verhoeven, 2005; Georgiou, Torppa, Manolitsis, Lyytinen, & Parrila, 2012). Thus, the mediated effect from vocabulary also ceases to exist. In the English language context, phonological processing is a strong predictor of reading for a longer time (e.g. Gallagher, Frith, & Snowling, 2000; Torgesen, Wagner, & Rashotte, 1994). In line with this, our results for the early grades show differences only in reading comprehension, while the studies from the English language context report differences in various reading skills (Duff et al., 2015; Rescorla, 2002). In later grades, however, our results are similar to the findings in English language context studies (Rescorla, 2005, 2009). It seems that the transparency of the orthography affects the speed of decoding development (e.g. Ellis et al., 2004; Seymour et al., 2003) and the importance of phonological processing in the development. Consequently, the mediated impact of expressive and/or receptive vocabulary delay via phonological processing to reading skills may thus be broader and may be sustained longer in the less transparent orthographies.

The limitations in this study include its small sample size and the inclusion of only one reading comprehension measure at each age. Analyses with a larger sample size would have more statistical power. However, samples with extensive assessments and such a long follow-up are rare and provide unique information. Further, it was not possible to identify a large enough group with difficulties in both expressive and receptive vocabulary but no family risk, which would have been informative. In addition, it would have been interesting to know how these children scored in expressive and receptive vocabulary at adolescence and whether the early language

difficulties experienced persisted and co-occurred with literacy skills in adolescence. Based on the findings by Rescorla (2005, 2009), we expect that this is the case, but unfortunately, we could not examine the issue due to the lack of such data in our sample. Moreover, data on family risk for reading comprehension or language skills would have been an interesting addition.

In conclusion, the findings indicate that the reading trajectories of late talkers depended on the co-occurrence of other risk factors and on the nature of the reading outcome in question. Early expressive vocabulary delay represented a high risk only for reading comprehension development and only when it co-occurred with other risk factors (family risk for dyslexia and receptive vocabulary difficulties). Late talking without receptive vocabulary difficulties did not increase the risk for reading difficulties among either group (family risk or no family risk). Our results indicate that in addition to expressive vocabulary delay, it is critical to recognize a delay in receptive skills as an early risk factor, since a delay in only expressive vocabulary is not informative enough to predict school age reading development. Based on the findings of the present study and those by Rescorla (2002, 2005, 2009), it is suggested that children with expressive vocabulary delay (especially those with family risk for dyslexia and receptive vocabulary delay as additional risk factors) should be provided extra support (e.g. exposure to language games and book reading). Finally, based on the results of the present study and those of previous studies, we suggest that the effects of receptive vocabulary should be examined more in depth, since all the studies so far have used a combination of expressive and receptive vocabulary delays. In addition, certain protective factors could lead to cognitive and socio-emotional resilience and influence the reading outcomes of children with reading difficulties. Based on the current findings, one potential protective cognitive skill may be receptive vocabulary. In line with the results of the study conducted by Haft, Myers and Hoeft (2016), we propose that future studies include investigations of the factors that contribute to the resilience of students who have reading difficulties or who are at risk of developing them.

Funding

Funding for this research was received from funding agencies in the public and not-for-profit sectors over the years. The major funder has been the Academy of Finland (current grants #276239, #284439, and #313768).

References

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

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

Aro, M. (2016). In L. T. W. Verhoeven, & C. A. Perfetti (Eds.). *Reading acquisition across languages and writing systems: An international handbook*. Cambridge: University Press.

Asikainen, M. (2005). *Diagnosing specific language impairment.* Finland: Academic dissertation, University of Tampere, Acta Electronica Universitatis Tamperensis.

Bayley, N. (1993). *The bayley scales of infant development* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech Language and Hearing Research, 49*(2), 278–293. http://dx.doi.org/10.1044/1092-4388(2006/023).

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*(8), 1106–1116. http://dx.doi.org/10.1177/0956797610375449.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848–856. http://dx.doi.org/10.1111/jcpp.12378.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250–287. http://dx.doi.org/10.1598/RRQ.36.3.2.

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology, 107*(1), 126–140. http://dx.doi.org/10.1037/a0037121.

Ellis, N. C., Natsume, I., Stavropoulou, K., Hoxhallari, L., van Daal, V. H. P., Polyzoe, N., et al. (2004). The effects of the orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*(4), 438–468. http://dx.doi.org/10.1598/RRQ.39.4.5.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856–868. http://dx.doi.org/10.1111/j.1467-8624.2005.00882.x.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* [(5, Ser. no. 242), i–185.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280–290. http://dx.doi.org/10.1111/jcpp.12010.

Gallagher, A., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry, 41*(2), 203–213. http://dx.doi.org/10.1111/1469-7610.00601.

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology, 100*(3), 566–580. http://dx.doi.org/10.1037/0022-0663.100.3.566.

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing: An Interdisciplinary Journal, 25*(2), 321–346. http://dx.doi.org/10.1007/s11145-010-9271-x.

Gough, P. B., & Tumner, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. http://dx.doi.org/10.1177/074193258600700104.

Häyrinen, T., Serenius-Sirve, S., & Korkman, M. (1999). *Lukilasse. Lukemisen, kirjoittamisen ja laskemisen seulontatesti ala-asteen luokille 1–6. Screening test for reading, spelling and counting for the grades 1–6*. Helsinki: Psykologien Kustannus Oy.

Haft, S. L., Myers, C. A., & Hoeft, F. (2016). Socio-emotional and cognitive resilience in children with reading disabilities. *Current Opinion in Behavioral Sciences, 10*, 133–141. http://dx.doi.org/10.1016/j.cobeha.2016.06.005.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*(2), 127–160. http://dx.doi.org/10.1007/BF00401799.

Hulme, C., & Snowling, M. J. (2014). The interface between spoken and written language: Developmental disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 20120395. http://dx.doi.org/10.1098/rstb.2012.0395.

Leinonen, S., Müller, K., Leppänen, P. H. T., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing: An Interdisciplinary Journal, 14*(3–4), 265–296.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study.

Developmental Psychology, 45(3), 764–781. http://dx.doi.org/10.1037/a0014132.
Lindeman, J. (2000). *Ala-asteen lukutesti: Tekniset tiedot ([2. p.]). ALLU Reading test for primary school: Technical information*. Turku, Finland: University of Turku, Centre for Research on Learning.
Lyytinen, P., Eklund, K., & Lyytinen, H. (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia, 55*(2), 166–192. http://dx.doi.org/10.1007/s11881-005-0010-y.
Lyytinen, H., Aro, M., Richardson, U., Erskine, J., Banff, A., Li, U. H., & Shu, H. (2015). Reading skills, acquisition of: Cultural, environmental, and developmental impediments. *International Encyclopedia of the Social & Behavioral Sciences (Second Edition),* 5–11. http://dx.doi.org/10.1016/B978-0-08-097086-8.23111-5.
Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*(2), 322–352.
Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language & Hearing Research, 50*(2), 508–528. http://dx.doi.org/10.1044/1092-4388(2007/035).
Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. http://dx.doi.org/10.1111/j.1467-9817.2004.00238.x.
Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031–1039. http://dx.doi.org/10.1111/j.1469-7610.2010.02254.x.
OECD (2010). *PISA 2009 results: Learning trends: Changes in student performance since 2000, vol. V*.
OECD (2013). *PISA 2012 results: What students know and can do: Student performance in mathematics, reading and science, vol. I*.
Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*(2), 189–208. http://dx.doi.org/10.1007/s11145-008-9159-1.
Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. http://dx.doi.org/10.1037/0022-0663.98.3.554.
Poll, G. H., & Miller, C. A. (2013). Late talking, typical talking, and weak language skills at middle childhood. *Learning and Individual Differences, 26*, 177–184. http://dx.doi.org/10.1016/j.lindif.2013.01.008.
Raven, J. C., Court, J. H., & Raven, J. (1992). *Standard progressive matrices.* Oxford, UK: Oxford Psychologists Press.
Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., ... Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics, 126*(6), e1530–e1537. http://dx.doi.org/10.1542/peds.2010-0254.
Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language & Hearing Research, 45*(2), 360–371. http://dx.doi.org/10.1044/1092-4388(2002/028).
Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language & Hearing Research, 48*(2), 459–472. http://dx.doi.org/10.1044/1092-4388(2005/031).
Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language & Hearing Research, 52*(1), 16–30. http://dx.doi.org/10.1044/1092-4388(2008/07-0171).
Reynell, J. K., & Huntley, M. (1987). *Reynell developmental language scales manual* (4th ed.). Windsor, UK: NFER-Nelson.
Ricketts, J., Nation, K., & Bishop, D. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235–257. http://dx.doi.org/10.1080/10888430701344306.
Rose, L. T., & Rouhani, P. (2012). Influence of verbal working memory depends on vocabulary: Oral reading fluency in adolescents with dyslexia. *Mind, Brain, and Education, 6*(1), 1–9. http://dx.doi.org/10.1111/j.1751-228X.2011.01135.x.
Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174. http://dx.doi.org/10.1348/000712603321661859.
Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an outlier orthography. *Psychological Bulletin, 134*(4), 584–615. http://dx.doi.org/10.1037/0033-2909.134.4.584.
Snowling, M. J., & Melby-Lervåg, M. (2016). Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychological Bulletin, 142*(5), 498–545. http://dx.doi.org/10.1037/bul0000037.
Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skill. *Child Development, 74*(2), 358–373. http://dx.doi.org/10.1111/1467-8624.7402003.
Snowling, M. J., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: A follow-up in early adolescence. *Journal of Child Psychology and Psychiatry, 48*(6), 609–618. http://dx.doi.org/10.1111/j.1469-7610.2006.01725.x.
Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language & Hearing Research, 48*(6), 1481–1495. http://dx.doi.org/10.1044/1092-4388(2005/103).
Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*(5), 276–286. http://dx.doi.org/10.1177/002221949402700503.
Torppa, M., Poikkeus, A., Laakso, M., Tolvanen, A., Leskinen, E., Leppanen, P. H. T., & Lyytinen, H. (2007). Modeling the early paths of phonological awareness and factors supporting its development in children with and without familial risk of dyslexia. *Scientific Studies of Reading, 11*(2), 73–103. http://dx.doi.org/10.1080/10888430709336554.
Torppa, M., Lyytinen, P., Erskine, J., Eklund, K., & Lyytinen, H. (2010). Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of Learning Disabilities, 43*(4), 308–321. http://dx.doi.org/10.1177/0022219410369096.
Walley, A. C., Metsala, J. L., & Garlock, V. M. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing, 16*(1–2), 5–20. http://dx.doi.org/10.1023/A:1021789804977.
van Bergen, E., de Jong, P. F., Plakas, A., Maassen, B., & van der Leij, A. (2012). Child and parental literacy levels within families with a history of dyslexia. *Journal of Child Psychology and Psychiatry, 53*(1), 28–36. http://dx.doi.org/10.1111/j.1469-7610.2011.02418.x.

II

LONGITUDINAL STABILITY OF READING DIFFICULTIES: EXAMINING THE EFFECTS OF MEASUREMENT ERROR, CUT-OFFS, AND BUFFER ZONES IN IDENTIFICATION

by

Maria Psyridou*, Asko Tolvanen, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus, & Minna Torppa, 2020

Frontiers in Psychology 10(2841)

frontiers in Psychology

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fpsyg.2019.02841

Longitudinal Stability of Reading Difficulties: Examining the Effects of Measurement Error, Cut-Offs, and Buffer Zones in Identification

Maria Psyridou[1], Asko Tolvanen[2], Marja-Kristiina Lerkkanen[1], Anna-Maija Poikkeus[1] and Minna Torppa[1]*

[1] Department of Teacher Education, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland,
[2] Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

This study examined the stability of reading difficulties (RD) from grades 2 to 6 and focused on the effects of measurement error and cut-off selection in the identification of RD and its stability with the use of simulations. It addressed methodological limitations of prior studies by (a) applying a model-based simulation analysis to examine the effects of measurement error and cut-offs in the identification of RD, (b) analyzing a non-English and larger sample, and (c) examining RD in both reading fluency and reading comprehension. Reading fluency and reading comprehension of 1,432 Finnish-speaking children were assessed in grades 2 and 6. In addition to the use of single cut-off points on observed data, we used a simulation approach based on an estimated structural equation model (SEM) in order to examine the effect of measurement error on RD identification stability. We also examined the effect of single cut-offs by using a simulation-based buffer zone. Our results showed that measurement error affects the identification of RD over time. The use of a simulation-based buffer zone could control both the effects of measurement error and the arbitrariness of single cut-offs and lead to more accurate classification into RD groups, especially for those with scores close to the cut-offs. However, even after controlling for measurement error and using buffer zones, RD was not stable over time for all children, but both resolving and late-emerging groups existed. The findings suggest that reading development needs to be followed closely beyond the early grades and that reading instruction should be planned according to individual needs at specific time points. There is a clear need for further consideration of the mechanisms underlying the stability and instability of RD.

Keywords: reading difficulties, stability, measurement error, cut-offs, simulation, reading fluency, reading comprehension

OPEN ACCESS

Edited by:
José Carlos Núñez,
Universidad de Oviedo Mieres, Spain

Reviewed by:
Rui Alexandre Alves,
University of Porto, Portugal
Celestino Rodríguez,
Universidad de Oviedo Mieres, Spain
Raquel Fidalgo,
Universidad de León, Spain

****Correspondence:***
Maria Psyridou
maria.m.psyridou@jyu.fi

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Psychology

Received: *20 September 2019*
Accepted: *02 December 2019*
Published: *10 January 2020*

Citation:
Psyridou M, Tolvanen A, Lerkkanen M-K, Poikkeus A-M and Torppa M (2020) Longitudinal Stability of Reading Difficulties: Examining the Effects of Measurement Error, Cut-Offs, and Buffer Zones in Identification. Front. Psychol. 10:2841. doi: 10.3389/fpsyg.2019.02841

INTRODUCTION

Studies on reading difficulties (RD) have mainly focused on reading development during the early grades, and long-term longitudinal follow-up studies are scarce. The customary assumption about the persistence of RD is contested by findings in recent studies on the stability of RD. These studies indicate that in addition to persistent RD cases, there are those with resolving RD (i.e., normal

reading skills in later grades despite RD being identified during early grades) and those with late-emerging RD (i.e., RD identified during the later grades despite normal early reading skills) (e.g., Catts et al., 2012; Torppa et al., 2015; Etmanskie et al., 2016). These unstable groups are of particular interest because they may provide more understanding on the developmental risk and supportive mechanisms in RD. Such knowledge is useful for more accurate identification of children with RD and for planning support for children and youth with RD. But do these groups truly exist, or are they simply a result of measurement error? When we categorize continuous reading distribution at two time points using arbitrary cut-offs, is it possible that the changes in RD classification are actually due to measurement error?

We cannot answer these questions based on the previous studies on RD stability because of methodological limitations. Such limitations include the use of (arbitrary) cut-offs in reading skill distributions to identify RD cases (Francis et al., 2005; Branum-Martin et al., 2013), not examining the role of measurement error played in the stability of diagnosis across time, small sample sizes (Leach et al., 2003; Lipka et al., 2006), and the use of very lenient 20–25% cut-offs in RD identification (e.g., Lipka et al., 2006; Etmanskie et al., 2016). Furthermore, nearly all studies except for Torppa et al. (2015) have been conducted in English, which constrains the generalization of the findings to other orthographies. Torppa et al. (2015), however, involved only a small number of Finnish children with RD from a specific sample with family risk for dyslexia and examined only reading fluency. Our study examines the effects of measurement error and cut-offs in the longitudinal stability of RD identification across two time points, grades 2 and 6, by using a structural equation model (SEM)-based simulation approach. Examining the effects of measurement error with the use of observed data is not possible because measurement error is constant in observed variables and cannot be manipulated. It can be manipulated though with the use of simulations (Schatschneider et al., 2016). In addition to the use of single cut-off, we examined the stability of RD identification by using a simulation-based buffer zone in order to handle both the effects of the use of single cut-off on continuous distributions (Shankweiler et al., 1999) and the effects of the measurement error. In our study, we used a large-population-based sample of Finnish-speaking children and included both reading fluency and reading comprehension.

Reading Fluency and Reading Comprehension Difficulties

One of the main aims of education is to teach young children to read and use reading for learning. To be considered a skilled reader, an individual must be able to read accurately and fluently and comprehend the meaning of what was read. Reading accuracy, fluency, and comprehension are closely linked skills, particularly in the early phases of reading development, and reading comprehension can only develop after some basic word identification skills have been achieved (e.g., Florit and Cain, 2011). According to the verbal efficiency theory (Perfetti, 1985), good reading fluency skills facilitate reading comprehension because automaticity in decoding reduces the resource demands of cognitive processes (e.g., memory and attention), which can then be allocated to comprehension. The strength of the relationship between reading fluency and reading comprehension diminishes over time (Florit and Cain, 2011; Torppa et al., 2016; Santos et al., 2019) as children become "fluent enough". This finding is in accordance with the Simple View of Reading (SVR) model (Gough and Tunmer, 1986; Hoover and Gough, 1990), which states that reading comprehension is a product of two separable abilities – decoding and linguistic comprehension. Several studies over the past 30 years have supported the SVR model (e.g., Catts et al., 2006; Tunmer and Chapman, 2012; Torppa et al., 2016; see García and Cain, 2014, for a meta-analysis). Based on SVR, it is thus expected that, although decoding and reading comprehension are highly related skills, four discrete groups of RD can be identified: only decoding difficulties, only reading comprehension difficulties, both decoding and reading comprehension difficulties, and no RD. Evidence for the dissociated reading skill development in decoding (accuracy and/or fluency) and reading comprehension has been shown repeatedly (e.g., Catts et al., 2003; Torppa et al., 2007; Florit and Cain, 2011).

As reading fluency and reading comprehension are associated, but difficulties with either can emerge; the examination of RD stability should include both. We base our RD stability analysis on a model that includes reading fluency and comprehension, their stability in time, their correlations at both time points, and cross-lagged effects. We do not include reading accuracy in our model because in transparent orthographies (Aro and Wimmer, 2003; Seymour et al., 2003; Aro, 2017) acquisition of high accuracy is a fast process and because almost all children read accurately after 1 year of formal instruction of reading (Lerkkanen et al., 2004; Landerl and Wimmer, 2008; Soodla et al., 2015).

RD Stability Over Time and the Challenge With Measurement Error in Cut-Off-Based RD Identification

Five studies have addressed the stability of RD classification over time (Leach et al., 2003; Lipka et al., 2006; Catts et al., 2012; Torppa et al., 2015; Etmanskie et al., 2016). Overall, these studies suggest that while some children have stable RD, many also show difficulties only in later grades despite good early reading skills (late-emerging RD). Some studies have also reported that there are individuals whose RD resolve over time (Catts et al., 2012; Torppa et al., 2015). One of the most cited studies of late-emerging poor readers was conducted by Leach et al. (2003), in which 161 grade 4 and grade 5 students were assessed based on school and parental reports. As a cut-off, they used a combination of the research-based estimation that 10–20% of children have RD and the range of reading abilities (including both word reading and reading comprehension measures) of the group with early school-identified and persistent RD. Thirty-one children with late-emerging RD were identified: 11 with word RD, 10 with reading comprehension difficulties, and 10 with both word reading and reading comprehension difficulties. In another longitudinal study, Lipka et al. (2006) examined word reading among 44 children with early RD, of which 22 had RD in grade 4

and 22 were typical readers. Using the 25th percentile as a cut-off in a standardized oral reading test, three subgroups of RD were identified: persistent RD (N = 7), borderline RD (N = 7), and late-emerging RD (N = 8). However, in their studies Leach et al. (2003); Lipka et al. (2006), and Torppa et al. (2015) used single cut-off points on raw variables with measurement error included. Consequently, it is possible to include false-positive and false-negative cases, which can affect the reliability of the results and the prevalence of each group.

Setting a cut-off value in reading ability distribution is a practical tool for identifying RD. However, it causes uncertainty in research findings because of measurement error (Francis et al., 2005; Branum-Martin et al., 2013; Schatschneider et al., 2016). When we use a reading test, there is going to be some measurement error, and as a result, setting a cut-off based on the raw scores will lead to misclassifications. If an individual's score is slightly above the cut-off, the measurement error could cause his/her observed score to fall below the cut-off, leading him/her to be falsely identified as having RD. Measurement error can thus affect the accurate identification of children with RD. It can also affect the stability of RD, because changes in RD status can reflect either a true change or be due to the effect of measurement error at either or both time points.

The effects of the use of cut-off points on raw variables with measurement error included were tested in the only non-English RD stability study (Torppa et al., 2015). In this Finnish study, the stability of RD status was examined from grades 2 to 8. In the study, 182 children participated, of which 101 had family risk for RD and 81 had no risk. Three reading speed tasks were used for the identification of RD in grades 2 and 8, and the 10th percentile was used as a cut-off point in the distribution of the children without family risk. Four groups were identified: no RD (N = 127), late-emerging RD (N = 18), resolving RD (N = 15), and persistent RD (N = 22). In addition to the cut-off-based identification of RD groups, a simulation approach was used to examine how many children would have changed their group due to the unreliability of the measurement. The simulation results showed that 10 out of 33 RD children were misclassified. While the simulation did confirm that RD seem not to be stable for all participants, it also showed that measurement error had a clear effect.

A different methodology was used by Catts et al. (2012), who examined the prevalence of late-emerging RD with a form of latent transition analysis (LTA) in a sample of 493 children followed through grades 2, 4, 8, and 10. Importantly, they examined RD in both word reading and reading comprehension. The LTA provides a good solution for modeling the transitions between latent classes over time. At each time point, four latent classes were allowed in the model: typical reader, word RD, reading comprehension difficulties, and both word reading and reading comprehension difficulties. The cut-off of 1 standard deviation below the weighted sample mean was used for the identification of RD at each time point. In their analysis, six groups were identified across time, two stable over time, and four with transitions. This study controlled measurement error in the LTA mover–stayer model by using multiple binary indicators. Although measurement error was controlled, the actual effects on transitions were not addressed. In addition, although all possible combinations for late-emerging RD were examined (late-emerging word RD, late-emerging reading comprehension difficulties, late-emerging word RD and reading comprehension difficulties), the persistent and resolving classes were not examined.

It is possible though that the single cut-offs contribute to false impressions of the distinctness of the RD groups if many resolving and late-emerging individuals are scoring just above the cut-off value. Shankweiler et al. (1999) argued that using a buffer zone instead of simple cut-off points divides better those with and those without RD. The latest edition of the American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) includes severity ratings, which reflect the idea of continuous reading distribution. Developmental disorders are the results of many risk factors and are better seen as dimensional disorders rather than diagnostic categories (Pennington, 2006; Snowling and Hulme, 2012; van Bergen et al., 2014). This is also in line with the notion of the arbitrariness of cut-off points for the classification into learning disorder groups (Moll et al., 2014). Etmanskie et al. (2016) used a buffer zone in their examination of the prevalence and persistence of late-emerging and early identified RD in a sample of 964 children. They used the 25th percentile as a cut-off for the identification of RD (word reading and/or reading comprehension), but to be considered a typical reader, a child's score needed to be at or above the 35th percentile (a buffer zone between the 25th and 35th percentiles). In grade 4, five groups were identified: typical reading skills (N = 694), word RD (N = 7), reading comprehension difficulties (N = 121), word reading and reading comprehension difficulties (N = 24), and borderline reading skills (N = 118). The children with poor word reading and/or reading comprehension difficulties were further regrouped into early identified, late-emerging, and inconsistent readers based on their performance in grades 1, 2, and 3.

The Present Study

The aim of the current study is to examine the effects of measurement error and the effects of single cut-offs in the stability of RD identification from grades 2 to 6. The study aims to address the methodological limitations of the previous studies on RD stability by (a) applying a model-based simulation analysis to examine the effects of measurement error and single cut-offs in the identification of RD, (b) analyzing a non-English and larger sample relative to previous ones, and (c) examining RD in both reading fluency and reading comprehension. The advantages of the use of simulation are threefold: we can examine the effects of measurement error and the effects of single cut-offs on transitions; we can examine all possible combinations for reading fluency and reading comprehension between grades 2 and 6 for all the groups (persistent RD, late-emerging RD, resolving RD); and our larger sample allows us to identify more groups that probably would not have been identified in observed data because some of these groups might not have been big enough. In addition, in our study, we focus on the beginning and the end of primary school in Finland. Because in grade 1 (during which formal instruction of reading begins), it would

be difficult to assess accurately reading comprehension, our first assessment point was grade 2. Grade 6 is the last grade of primary school. After that, children enter junior high school and high school where they are taught by subject teachers instead of the classroom teacher.

The research questions of the present study are as follows:

1. How stable are reading fluency and reading comprehension RD from grades 2 to 6?
2. What is the effect of measurement error on the estimation of RD stability over time from grades 2 to 6?
3. What is the effect of using single cut-offs compared with a buffer zone when examining RD stability over time from grades 2 to 6?

We expect that cases of late-emerging and resolving RD would also emerge in the present data (e.g., Leach et al., 2003; Catts et al., 2012; Torppa et al., 2015; Etmanskie et al., 2016). We also anticipated that the simulation approach would reveal an effect of measurement error (Schatschneider et al., 2016) and that the use of a buffer zone would lower the percentages of changing RD groups (resolving and late-emerging).

MATERIALS AND METHODS

Participants and Procedure

The present study is part of the longitudinal First Steps Study (Lerkkanen et al., 2006), a follow-up of approximately 2,000 children from kindergarten to grade 6. The aim of the First Steps Study is to examine the development of children's learning and motivation in the family and school contexts. The sample was drawn from four municipalities: two in central, one in western, and one in eastern Finland. In three of the municipalities, an invitation was sent through schools to the whole age cohort of children, and in the fourth (urban) municipality, the invitation for participation was sent to approximately half of the age cohort. At the beginning of the study, the children's parents and teachers were asked for written consent. Of the parents who were contacted, 78–89%, depending on town or municipality, agreed to take part in the study. Of the children's mothers, 7.6% had no education beyond secondary school, 30.2% had a vocational school degree, 23.8% a vocational college degree, 9.9% a bachelor's degree, 24% a master's degree, and 4.6% a doctoral degree. Of the children's fathers, 7.9% had no education beyond secondary school, 33.2% had a vocational school degree, 23.7% a vocational college degree, 9.9% a bachelor's degree, 19% a master's degree, and 6.3% a doctoral degree. Parental education distribution was very close to the national distribution of Finland (Statistics Finland, 2007). The sample was highly homogeneous in ethnic and cultural background (e.g., Finnish-speaking schools and students). The study was approved by the Ethical Committee of the University of Jyväskylä and at the beginning of the study the children's parents, and teachers provided informed written consents for participation. During the study, also the children gave their written consent to participate.

The present study involved assessments at two time points – the end of grade 2 (Spring 2009) (8 years) and the end of grade 6 (Spring 2013) (13 years). Only children (N = 1,432; 662 girls and 770 boys) for whom data were available for both the grade 2 and grade 6 assessments were included in the analyses. All participants who were assessed in grade 6 were included in the current sample. In grade 6 spring, 1,824 12- to 13-year-old children participated: 863 girls (47.31%) and 961 (52.69%) boys. Of them, 1,458 participated also in grade 2 spring (8–9 years old): 680 girls (46.64%) and 778 (53.36%) boys. The sample size of the First Steps Study changed somewhat each year due to factors as shifts in teaching groups or absences during the testing days. In the present study, data from 72 schools and 147 classrooms were used. The SEM described below for the development of reading fluency and reading comprehension from grades 2 to 6 was also constructed using full grade 6 data. There were only minor differences to some of the path estimates ranging from 0.01 to 0.05.

Conducting such a long-term follow-up study is challenging, and some changes in the sample from one assessment time point to another are inevitable. We conducted a missing value analysis in order to examine if missingness was random for the data we used at the current study (z-scores for reading fluency composite and reading comprehension in grades 2 and 6). We used the Little's (1988) tests of missing completely at random (MCAR), which showed that the data were not MCAR, $\chi^2(14) = 62.29$, $p < 0.001$. Reading fluency composite score and reading comprehension in grade 2 had 20.34 and 21.27% of the cases missing. Reading fluency composite score and reading comprehension in grade 6 had 0.22 and 0.16% of the cases missing. The one-way ANOVA analysis comparing the reading fluency and reading comprehension performance of the sample included in this study (only those who were assessed in both grades 2 and 6; N = 1,432) and the whole sample (N = 1,824) showed that those who participated in both grades were somewhat better readers than those who participated only in grade 6 [for reading fluency: $F(1,1,812) = 47.08$, $p < 0.001$; for reading comprehension: $F(1,1,819) = 7.46$, $p < 0.01$]. However, the effect sizes were small for reading fluency ($d = -0.39$) and negligible for reading comprehension ($d = -0.15$).

The comprehensive education, grades 1–9, starts at the fall from the year in which the child turns 7 years of age, which is rather late compared with other countries. Before entrance to elementary school, all 6-year-olds attend 1-year kindergarten education. One goal of kindergarten education is to arouse children's interest in texts and reading and to support emerging pre-reading skills, instead of a systematic instruction of decoding (Lerkkanen, 2018). However, children are read to, and they are also encouraged to play with letters, phonemes, and words (Lerkkanen, 2019). Reading instruction begins at grade 1, and it is based on grapheme–phoneme correspondence (phonics) and a highly transparent Finnish orthography, which makes reading acquisition relatively easy and quick for children (Lerkkanen, 2007). At the end of kindergarten, around 30% of the children can read fluently, around 30% can decode easy words while around 30% of the children show no sign of reading (Lerkkanen et al., 2010; Soodla et al., 2015). Largely due to the consistent nature of the highly transparent orthography of the Finnish language (Aro, 2017), reading accuracy hits a ceiling

after a few months of formal reading instruction in grade 1 (Lerkkanen et al., 2004), and basically all children can read accurately by the end of the first school year (Soodla et al., 2015). However, even a highly consistent orthography does not guarantee efficient reading acquisition for all children. RD are typically identified for approximately 5–20% of children in either reading fluency or comprehension, depending on the criteria (Lerkkanen et al., 2010).

In basic education, children do not need to have an official diagnosis in order to have access to special educational services. Teachers and parents along with the students assess the need for extra support (Björn et al., 2016). The most common form of special educational services is the part-time special education provided by a special education teacher (Statistics Finland, 2005). In this form of special education, students study in general education classes and receive support 1–2 h/week from a special education teacher. This kind of support focuses on reading, spelling, and math difficulties. It is implemented in small groups (typically three to four students) or individually if the student faces long-lasting or more severe difficulties or if the student faces difficulties in more than one learning areas (Holopainen et al., 2018).

Measures

Reading Fluency

There were three group-administered tests for the assessment of reading fluency: a word reading fluency task, a word-chain task, and a sentence reading task. Cronbach's alpha reliability coefficient for the fluency composite was 0.79 in grade 2 and 0.77 in grade 6.

Word reading fluency task

The word reading fluency task is a subtest of the nationally normed reading test battery [ALLU–Ala-asteen lukutesti (ALLU–Reading Test for Primary School); Lindeman, 2000]. Each of the 80 items consisted of a picture with four phonologically similar words attached to it. The child silently read the four words and then drew a line to connect the picture with the word, semantically matching it. The words and pictures were frequently used words familiar to young children. For example, there was a picture of a bunny (pupu in Finnish) and the correct word along with three distractors (English word is in parentheses): pipo (cap), papu (bean), and apu (help). Completing the test requires fluent decoding. The score was the number of correct answers within a 2-min time limit. Because of the nature of this timed test, the score reflects both the child's fluency in reading the stimulus words and accuracy in making the correct choice from among the alternatives. According to the test manual (Lindeman, 1998), the Kuder–Richardson reliability was 0.82 in grade 2 and 0.97 in grade 6.

Word-chain task

The word-chain task (Nevala and Lyytinen, 2000) is a timed test with 10 rows of word chains comprising four to six words written together without spaces. The child silently read the words in the chains and, while reading them, indicated the word boundaries by drawing a division line between words. The score was the number of correct responses (maximum 40) within the time limit (1 min 25 s in grade 2). In our sample, the Pearson correlation coefficient between grades 2 and 6 was 0.52.

Sentence reading task

The Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al., 2009; Finnish version by Lerkkanen and Poikkeus, 2009) was used to assess silent reading efficiency in grade 2. Respondents were given 3 min to read 60 sentences and verify the truthfulness of as many sentences as possible. In grade 6, a similar task was used, the Salzburg Sentence Reading Test (Landerl et al., 1997, translated into Finnish by Sini Huemer; Pichler and Wimmer, 2006). Respondents were given 2 min to read 69 sentences and verify the truthfulness of as many sentences as possible. The sum score was based on the number of correct items. In our sample, the Pearson correlation coefficient between grades 2 and 6 was 0.67.

Reading Comprehension

A group-administered subtest of a nationally normed reading test battery (ALLU; Lindeman, 2000) was used to assess reading comprehension. The children silently read a fiction story and then answered 11 multiple-choice questions and one question in which they had to arrange five statements in the correct sequence based on information gathered from the text. The text contained 114 words in grade 2. The child received 1 point for each correct answer (max = 12). Each child completed the task at his or her own pace, but the maximum time allotted was 45 min. Lindeman (2000) reported Kuder–Richardson reliability coefficients of 0.80 in grade 2 and 0.74 in grade 6.

Analysis Description

First, we identified the RD groups using cut-off points on the observed data in a similar fashion as previous RD stability studies. We first calculated z-scores for reading comprehension and z-scores for the three reading fluency tasks in grades 2 and 6 separately. Based on the z-scores for reading fluency, we calculated mean composite scores for each grade. We used the 10th percentile as the cut-off value and dichotomized the reading fluency and reading comprehension z-score variables accordingly. The four variables were coded at each time point as 0 = typical reader (above the 10th percentile) and 1 = RD (below the 10th percentile) for each case (**Table 3**).

The simulation analysis on the effects of measurement error on the RD grouping started by building a SEM for the development of reading fluency and reading comprehension from grades 2 to 6 (see **Supplementary Appendix A**). The model was constructed using four latent variables consisting of separate factors for reading fluency and reading comprehension in grades 2 and 6 (see **Figure 1**). The use of latent variables for reading fluency (composed of three measures) and reading comprehension provide reading measures that do not include measurement error. Because reading comprehension had only one measure at each time point, we calculated the correction of attenuation using the Kuder–Richardson reliability estimates for reading comprehension in each grade from the test manual (Lindeman, 2000). In this way, we can set measurement error also

TABLE 1 | Descriptive statistics for reading fluency and reading comprehension measures in grades 2 and 6.

	N	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
Grade 2							
Word reading fluency task	1,458	3	58	24.75	7.39	0.46	0.11
Word-chain task	1,458	0	35	11.59	5.93	0.61	0.31
Sentence reading task	1,453	3	60	30.57	8.15	0.02	0.26
Reading comprehension	1,436	0	12	8.73	2.58	−0.79	−0.14
Grade 6							
Word reading fluency task	1,820	10	80	47.22	10.94	0.00	0.00
Word-chain task	1,820	1	40	21.63	7.44	0.09	−0.36
Sentence reading task	1,822	4	62	30.61	7.38	0.15	0.28
Reading comprehension	1,821	0	12	7.15	2.55	−0.20	−0.59

in reading comprehension. The model included the stability paths within the reading constructs, the cross-lagged paths between reading fluency and comprehension across time, and correlations between reading fluency and comprehension at both time points.

The structural equation model analyses were carried out using Mplus 7.4 software. Maximum likelihood estimation with robust standard errors (MLR) was used for the analysis. To evaluate model fit, chi-square values and a set of fit indexes were used as follows: (a) the comparative fit index (CFI); (b) Tucker–Lewis index (TLI); and (c) the root-mean-square error of approximation (RMSEA). Good model fit is indicated by a small, preferably non-significant χ^2, CFI > 0.95, TLI > 0.95, and RMSEA < 0.06 (Hu and Bentler, 1999). Since the chi-square test depends on sample size and is sensitive to a large sample size, the chi-square statistics were not regarded as conclusive.

Next, we estimated model parameters of the SEM model to produce two simulated datasets with 200,000 cases. The first dataset was simulated using parameters related to latent factors, hence corresponding to true scores without measurement error (see **Supplementary Appendix B**). The second dataset used all the parameters in the model to produce data that include also the measurement error (see **Supplementary Appendix C**). In this dataset, the error covariances were also included in the simulation equations to correspond to the observed situation. The simulations were produced using the Statistical Package for the Social Sciences (SPSS). We used the estimates for the stability paths, the cross-lagged paths, and the correlations between reading fluency and reading comprehension variables. For each simulation, we entered four discrete variables into the model: fluency in grade 2, fluency in grade 6, comprehension in grade 2, and comprehension in grade 6. The scores for each of the four variables were coded at each time point as 0 = typical reader (highest 90%) and 1 = RD (lowest 10%) for each case. Next, we calculated the frequencies of RD groups in each simulated dataset. By comparing the percentages of RD groups in the simulated samples without and with measurement error, we can examine the effect of measurement error on RD identification and RD identification stability.

The final step of our analysis was to examine RD stability with the use of a buffer zone. We used the same procedure described above, but the four discrete variables that were entered into the model were coded at each time point as 0 = typical reader (highest 75%), 1 = RD (lowest 10%), and 2 = borderline score (lowest 10–25%) for each case. Those with scores in the lowest 10% of the reading fluency and/or reading comprehension distribution were identified as manifesting reading fluency and/or reading comprehension difficulties. Those with scores in the lowest 10–25% range were identified as borderline readers, and those with scores above 25% were identified as having no RD (**Tables 7**, **8**). In this analysis, we focused specifically on the late-emerging and the resolving RD in order to examine how far the scores are from the cut-off and how distinct the two groups are that were identified using the single cut-off.

RESULTS

Descriptive Statistics

See **Table 1** for descriptive statistics and **Table 2** for the correlations between reading fluency and reading comprehension measures in grades 2 and 6. For the reading measures, the stability correlations between grades 2 and 6 ranged from moderate to high (0.49–0.67).

Identification of the RD Groups With Observed Data

First, RD groups in reading fluency and reading comprehension in grades 2 and 6 were calculated from observed data using the 10th percentile as the cut-off value (**Table 3**). There was a

TABLE 2 | Correlations between reading fluency and reading comprehension measures.

	1	2	3	4	5	6	7	8
Grade 2								
(1) Word reading fluency task	1							
(2) Word-chain task	0.48	1						
(3) Sentence reading task	0.66	0.53	1					
(4) Reading comprehension	0.31	0.40	0.43	1				
Grade 6								
(5) Word reading fluency task	0.58	0.38	0.56	0.31	1			
(6) Word-chain task	0.43	0.52	0.53	0.39	0.52	1		
(7) Sentence reading task	0.53	0.44	0.67	0.45	0.63	0.52	1	
(8) Reading comprehension	0.23	0.29	0.33	0.49	0.22	0.32	0.38	1

statistically significant association between RD in grade 2 and RD in grade 6 [$\chi^2(9) = 441.71$, $p < 0.001$].

Approximately 10% of the participants had persistent RD, most of whom had persistent single reading fluency or reading comprehension difficulties. Most of those with either reading fluency or reading comprehension difficulties in grade 2 continued to have the same difficulty in grade 6, or they developed both reading fluency and comprehension difficulties in grade 6. Most of those with both reading fluency and reading comprehension difficulties in grade 2 continued to have only reading fluency difficulties or both reading fluency and reading comprehension difficulties in grade 6. It was rare for those struggling with any type of RD in grade 2 to move to no RD group in grade 6 and the other way around.

The percentage of children with no RD in grade 2 but some type of RD in grade 6 (late-emerging cases) was quite high (12.71%), and two-thirds of these cases had late-emerging reading comprehension difficulties. Only 2.79% of the children had only late-emerging reading fluency difficulties. The percentage of resolving difficulties (some kind of RD in grade 2 but no RD in grade 6) was smaller (6.42%), and the prevalence of resolving reading fluency and resolving reading comprehension difficulties was quite similar. Both resolving and late-emerging profiles were rare among children with both reading fluency RD and comprehension RD.

In order to describe the severity of RD in each group, we calculated means and standard deviations of the RD groups in reading fluency and reading comprehension in grades 2 and 6 (**Table 4**). Those with persistent RD performed 1.5 or more standard deviations below the average level in either fluency, comprehension, or both. Similarly, those with resolving RD performed 1.5 or more standard deviations below the average in grade 2, and those with late-emerging RD performed 1.5 or more standard deviations below the average in grade 6. However, in the resolving RD groups, the grade 6 skill performance was still below average, particularly for reading fluency (as low as −0.88 in the combined group). Similarly, the grade 2 reading levels of late-emerging RD groups were also somewhat below average (as low as −0.76 in the combined group).

Gender was unevenly distributed in the groups, $\chi^2(3) = 15.32$, $p < 0.05$: there were less boys than expected in the group with no RD in both grades (68.38% of boys compared with 77.80% of girls with adjusted standardized residual = −3.43) (**Table 5**). In the RD groups, there were no significant differences. Of the boys, 11.55% belonged to the persistent RD groups, 14.39% to the late-emerging group, and 5.68% to the resolving group with adjusted standardized residuals of 2.69, 2.50, and −0.29, respectively. Among the girls, 6.72% belonged to the persistent RD groups, 9.37% to the late-emerging groups, and 6.11% to the resolving groups.

Simulation-Based Identification of the RD Groups: The Effect of Measurement Error

In order to examine whether measurement error affects the identification and stability of RD across time, we produced one simulation without and another with measurement error.

TABLE 3 | Number and percentage of children in each group based on the observed data for those who were assessed in both grades 2 and 6 using the 10th percentile as a cut-off point.

			Grade 6				
			RF only	**RC only**	**RF + RC**	**no RD**	**Total**
Grade 2	RF[a] only	Number of children	35	17	9	43	104
		% within grade 2	33.65	16.35	8.65	41.35	100.00
		% of total	2.44	1.19	0.63	3.00	7.26
		Adjusted residual	11.58	1.26	3.05	−9.20	
	RC[b] only	Number of children	7	30	13	38	88
		% within grade 2	7.95	34.09	14.77	43.18	100.00
		% of total	0.49	2.09	0.91	2.65	6.15
		Adjusted residual	0.54	6.36	6.05	−7.99	
	RF + RC	Number of children	12	2	14	11	39
		% within grade 2	30.77	5.13	35.90	28.21	100.00
		% of total	0.84	0.14	0.98	0.77	2.72
		Adjusted residual	6.19	−1.40	11.31	−7.50	
	no RD[c]	Number of children	40	129	13	1,019	1,201
		% within grade 2	3.33	10.74	1.08	84.85	100.00
		% of total	2.79	9.01	0.91	71.16	83.87
		Adjusted residual	−11.27	−4.42	−11.10	15.03	
Total		Count	94	178	49	1,111	1,432
		% within grade 2	6.56	12.43	3.42	77.58	100.00
		% of total	6.56	12.43	3.42	77.58	100.00

[a]*RF, reading fluency;* [b]*RC, reading comprehension;* [c]*no RD, no reading difficulties.*

TABLE 4 | Descriptive statistics for the groups based on the observed data.

	N	Reading fluency, grade 2^z		Reading comprehension, grade 2^z		Reading fluency, grade 6^z		Reading comprehension, grade 6^z	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Persistent									
RC[a] → RC	30	−0.37	0.60	−2.15	0.31	−0.37	0.70	−1.58	0.35
RC → RF[b]	7	−0.78	0.41	−1.99	0.30	−1.49	0.16	−0.71	0.31
RC → RF + RC	13	−0.85	0.24	−1.92	0.17	−1.66	0.21	−1.60	0.35
RF → RC	17	−1.47	0.19	−0.78	0.64	−0.63	0.34	−1.54	0.33
RF → RF	35	−1.58	0.33	−0.34	0.66	−1.75	0.40	−0.14	0.78
RF → RF + RC	9	−1.81	0.40	−0.67	0.64	−2.04	0.38	−1.74	0.38
RF + RC → RC	2	−1.66	0.38	−1.83	0.00	−1.19	0.10	−1.66	0.00
RF + RC → RF	12	−1.95	0.36	−2.37	0.42	−1.89	0.41	−0.29	0.56
RF + RC → RF + RC	14	−1.88	0.46	−2.24	0.44	−1.95	0.50	−1.88	0.40
Late emerging									
no RD → RC	129	−0.06	0.75	−0.21	0.79	−0.05	0.67	−1.54	0.36
no RD → RF	40	−0.64	0.47	−0.28	0.77	−1.56	0.19	−0.01	0.63
no RD → RF + RC	13	−0.76	0.32	−0.70	0.46	−1.60	0.22	−1.72	0.39
Resolving									
RC → no RD[c]	38	−0.28	0.54	−2.07	0.33	−0.24	0.63	−0.28	0.57
RF → no RD	43	−1.49	0.24	−0.22	0.74	−0.69	0.48	0.09	0.73
RF + RC → no RD	11	−1.55	0.22	−2.14	0.38	−0.88	0.27	−0.34	0.59
No RD									
no RD → no RD	1, 019	0.31	0.88	0.35	0.72	0.32	0.86	0.38	0.76
Total	1, 432								

Subscript z refers to standardized score. The reading fluency measures are composite scores calculated as sum scores of the standardized scores for the three reading fluency tasks. [a]RC, reading comprehension; [b]RF, reading fluency; [c]no RD, no reading difficulties. On the left side of the arrow is the reading difficulty status in grade 2 [only reading fluency difficulties (RF), only reading comprehension difficulties (RC), both reading fluency and reading comprehension difficulties (RF + RC), or no reading difficulties (no RD)]. On the right side of the arrow is the reading difficulty status in grade 6. Reading fluency and/or reading comprehension difficulties were calculated using the 10th percentile cut-off value.

SEM Model

Latent factors for reading fluency and reading comprehension were built in grades 2 and 6 (see **Figure 1**). In addition to the regression paths across time (grades 2 and 6) and across constructs (reading fluency and reading comprehension), residual covariances for each measure of reading fluency across time were added to the model based on the inspection of modification indices. The model showed good fit with the data, $\chi^2(14) = 117.163$, $p < 0.001$, RMSEA = 0.072, CFI = 0.977, TLI = 0.954, standardized root mean square (SRMR) = 0.032. The model indicated that reading fluency was very stable across time while reading comprehension was less stable. Of grade 6 reading fluency variance, 60.8% was explained by grade 2 reading fluency and an additional 2% by grade 2 reading comprehension. Of grade 6 reading comprehension variance, 32.5% was explained by grade 2 reading comprehension and an additional 1.2% by grade 2 reading fluency.

Next, based on the model, we produced a simulated dataset without measurement error. We simulated 200,000 cases and identified cases with RD as the lowest 10% of the reading fluency and/or reading comprehension distribution (**Table 6**). In the simulated data without measurement error, 86.45% were in the stable groups (76.52% no RD; 9.93% persistent RD) and 13.55% demonstrated instability in RD across grades. Of the cases demonstrating instability, 6.50% had resolving RD and 7.05% were identified as manifesting late-emerging RD; most of these manifested late-emerging reading comprehension difficulties.

Next, we produced a simulated dataset with 200,000 cases with measurement error. Overall, 82.53% of the cases were in stable groups from grades 2 to 6 (73.92% no RD; 8.61% persistent RD) (**Table 6**). The remaining 17.48% of the cases demonstrated instability in RD across grades. Of the cases demonstrating instability, 8.52% had resolving RD, and 8.96% were identified as manifesting late-emerging RD; most of these manifested difficulties in reading comprehension.

Finally, the group sizes produced by the two simulations were compared in order to examine the effects of measurement error. For the simulation without measurement error, the percentage of the stable groups was slightly higher, and the percentage of the cases with instability in RD was somewhat lower. The results, thus, suggest that 1.32% of the children would have not been identified as having persistent RD due to the inclusion of measurement error in the analysis. In other words, in the observed data of 1,432 children, approximately 19 children would be wrongly classified. Similarly, 3.93% of the children would have changed groups due to the inclusion of measurement error in the analysis, which means that in the observed data, 56 out of 1,432 children would be wrongly classified. More specifically, 28

TABLE 5 | Prevalence of boys and girls in each group based on observed data.

Group		Gender		Total
		Male	Female	
Persistent	Count	65	33	98
	% with the groups	66.33	33.67	100
	% within gender	11.55	6.72	9.30
	% of total	6.17	3.13	9.30
	Adjusted residual	2.69	−2.69	
Late emerging	Count	81	46	127
	% with the groups	63.78	36.22	100
	% within gender	14.39	9.37	12.05
	% of total	7.69	4.36	12.05
	Adjusted residual	2.50	−2.50	
Resolving	Count	32	30	62
	% with the groups	51.61	48.39	100
	% within gender	5.68	6.11	5.88
	% of total	3.04	2.85	5.88
	Adjusted residual	−0.29	0.29	
no RD	Count	385	382	767
	% with the groups	50.20	49.80	100
	% within gender	68.38	77.80	72.77
	% of total	36.53	36.24	72.77
	Adjusted residual	−3.43	3.43	
Total	Count	563	491	1,054
	% with the groups	53.42	46.58	100
	% within gender	100	100	100
	% of total	53.42	46.58	100

children (1.91%) would have been misclassified as having late-emerging RD and 28 children (2.02%) misclassified as having resolved their RD.

The Effect of the Use of a Single Cut-Off

The final step of our analysis was to examine RD stability with the use of a buffer zone. In this analysis, we focused particularly on the late-emerging RD (**Table 7**) and resolving RD groups (**Table 8**). Similar to the use of a single cut-off, we produced two simulated datasets, one without and one with measurement error. We simulated 200,000 cases for each dataset and identified RD for those cases located in the lowest 10% of the reading fluency and/or reading comprehension distribution. Cases with scores in the lowest 10–25% range were identified as having borderline scores, and cases with scores above 25% were identified as having no RD.

The results from the use of the buffer zone suggest that most of the cases from the late-emerging and resolving RD groups actually land in the buffer zone. For the late-emerging group, the simulation without measurement error and with the buffer zone showed that only 33 cases would be identified as having late-emerging RD compared with the 101 that were identified with the use of a single cut-off. Of the 33 cases, 28 were identified with late-emerging reading comprehension difficulties and five with late-emerging reading fluency difficulties (**Table 7**). For the resolving group, the simulation without measurement error and with the buffer zone showed that 28 cases would be identified as manifesting resolving RD compared with the 94 identified with the use of a single cut-off. Of the 28 cases, 19 would be identified with resolving reading comprehension difficulties, 8 with resolving reading fluency difficulties, and 1 with resolving reading fluency and comprehension difficulties (**Table 8**).

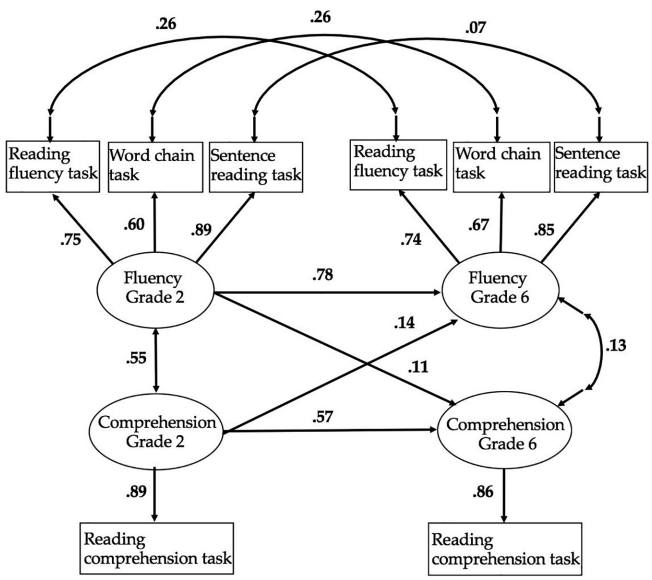

FIGURE 1 | Standardized path estimates of the cross-lagged model. The figure presents standardized estimates using listwise analysis.

TABLE 6 | Percentages and number of individuals in reading difficulty (RD) groups identified with the use of a single cut-off.

Group	Subgroup	Simulation without measurement error (group %)	Simulation with measurement error (group %)
Persistent	RC[a] → RC	1.92 (27)	1.57 (22)
	RC → RF[b]	0.54 (8)	0.73 (10)
	RC → RF + RC	0.47 (7)	0.44 (6)
	RF → RC	0.33 (5)	0.54 (8)
	RF → RF	2.93 (42)	2.56 (37)
	RF → RF + RC	0.68 (10)	0.61 (9)
	RF + RC → RC	0.36 (5)	0.35 (5)
	RF + RC → RF	1.28 (18)	1.02 (15)
	RF + RC → RF + RC	1.42 (20)	0.79 (11)
Persistent total		9.93 (142)	8.61 (123)
Late emerging	no RD[c] → RC	4.38 (63)	5.10 (73)
	no RD → RF	2.23 (32)	3.26 (47)
	no RD → RF + RC	0.44 (6)	0.60 (9)
Late-emerging total		7.05 (101)	8.96 (129)
Resolving	RC → no RD	3.50 (50)	4.38 (63)
	RF → no RD	2.48 (36)	3.41 (49)
	RF + RC → no RD	0.52 (8)	0.73 (10)
Resolving total		6.50 (94)	8.52 (122)
no RD	no RD	76.52 (1, 095)	73.92 (1, 058)

[a]RC, reading comprehension, [b]RF, reading fluency, [c]no RD, no reading difficulties. The number in parentheses shows the number of individuals based on the percentage for N = 1,432. On the left side of the arrow is the RD status in grade 2 [only reading fluency difficulties (RF), only reading comprehension difficulties (RC), both reading fluency and reading comprehension difficulties (RF + RC), or no reading difficulties (no RD)]. On the right side of the arrow is the RD status in grade 6. Reading fluency and/or reading comprehension difficulties were calculated using the lowest 10%.

DISCUSSION

The main focus of the present study was to examine the longitudinal stability of RD identification across two time points, grades 2 and 6 including both reading fluency and reading comprehension. We examined whether RD identification was stable over time even if we control for measurement error and the use of single cut-offs. Our results showed that for some children RD are not stable but also revealed a clear effect of measurement error and the selection of the cut-off in the identification of RD. The findings highlight that measurement error affects the accurate identification of children with RD by causing misclassifications and that the simplicity of the single cut-offs can contribute to false impressions on instability of RD over time.

All the previous studies, except the study conducted by Catts et al. (2012), examining the stability of RD identification, used cut-off points on raw variables with measurement error included (Leach et al., 2003; Lipka et al., 2006; Torppa et al., 2015; Etmanskie et al., 2016). This means that conclusions may be biased by false-positive or false-negative cases due to measurement error causing misclassifications. In this study, we used simulations with models that do and do not include measurement error to estimate the magnitude of this problem. Our findings comparing the results of the simulations with and without measurement error showed the impact of measurement error but still aligned with prior findings suggesting RD instability for a group of children (Leach et al., 2003; Lipka et al., 2006; Catts et al., 2012; Torppa et al., 2015; Etmanskie et al., 2016). In our study, 74–77% of participants were typical readers across time, and each RD group consisted of 7–10% of the participants (depending on the group and the model). Most of the participants (86.45% in the simulation without measurement error and 82.53% in the simulation with measurement error) demonstrated stability in their RD status (no RD and persistent RD). The simulation without measurement error revealed, however, larger proportions of the stable groups (persistent RD and no RD) and smaller proportions of late-emerging and resolving RD compared with the simulation with measurement error. These differences were expected because the results from the simulation with measurement error include false-positive or false-negative cases because of the effect measurement error, which affects the reliability of the estimation of the prevalence of each RD group. Although the differences in the prevalence of the groups were small, they show that measurement error has an effect on the longitudinal stability of RD identification. Overall, though, the findings supported that the unstable groups exist, even if we control for measurement error, which has been a problem in most previous studies.

Catts et al. (2012) study is the only previous study comparable to the present simulation analysis as they used LTA, which relied on multiple indicators for each reading class, and their findings were thus less affected by measurement error. Our results from the analyses using single cut-offs were in line with Catts et al. (2012) in that there was a higher prevalence of late-emerging reading comprehension than late-emerging reading fluency difficulties, but the findings show differences in the proportion of children in each RD group. More specifically, the present study identified a smaller proportion of late-emerging cases (7.05% compared with 13.40%) and more resolving cases (6.50% compared with 1.90%). Of the children with RD in the Catts et al. (2012) study, 42% had late-emerging RD and 6% had resolving RD; in our study, the percentages were 30 and 28%, respectively.

It is possible that differences in orthography, in assessment ages of children, or differences in the criteria used for the identification of RD, explain the differences between the present study and that by Catts et al. (2012). Finnish orthography is highly transparent, with one-on-one correspondence between phonemes and graphemes (Seymour et al., 2003; Lyytinen et al., 2015; Aro, 2017). Due to the transparency of Finnish orthography, most Finnish children learn to read accurately after

TABLE 7 | Percentage and number of individuals in late-emerging groups identified with the use of a buffer zone (bz).

Late-emerging groups identified using a single cut-off	Late-emerging groups identified using a buffer zone	Simulation without measurement error, group % (*N*)	Simulation with measurement error, group % (*N*)
no RD → RC	no RD → RC	1.69 (24)	2.30 (33)
	no RD → RF(bz) + RC	0.28 (4)	0.42 (6)
	RC(bz) → RC	1.10 (16)	0.98 (14)
	RC(bz) → RF(bz) + RC	0.29 (4)	0.32 (5)
	RF(bz) → RC	0.20 (3)	0.36 (5)
	RF(bz) → RF(bz) + RC	0.28 (4)	0.25 (4)
	RF(bz) + RC(bz) → RC	0.21 (3)	0.24 (3)
	RF(bz) + RC(bz) → RF(bz) + RC	0.33 (5)	0.22 (3)
		4.38 (63)	**5.09 (73)**
no RD → RF	no RD → RF	0.29 (4)	0.85 (12)
	no RD → RF + RC(bz)	0.09 (1)	0.22 (3)
	RC(bz) → RF	0.09 (1)	0.26 (4)
	RC(bz) → RF + RC(bz)	0.06 (1)	0.12 (2)
	RF(bz) → RF	0.82 (12)	0.92 (13)
	RF(bz) → RF + RC(bz)	0.26 (4)	0.31 (5)
	RF(bz) + RC(bz) → RF	0.37 (5)	0.37 (5)
	RF(bz) + RC(bz) → RF + RC(bz)	0.25 (4)	0.20 (3)
		2.23 (32)	**3.25 (47)**
no RD → RF + RC	no RD → RF + RC	0.04 (0)	0.13 (2)
	RC(bz) → RF + RC	0.06 (1)	0.11 (1)
	RF(bz) → RF + RC	0.14 (2)	0.18 (3)
	RF(bz) + RC(bz) → RF + RC	0.20 (3)	0.18 (3)
		0.44 (6)	**0.60 (9)**

The number in parentheses shows the number of individuals based on the percentage for N = 1,432. The bold values show the total number of late-emerging cases.

TABLE 8 | Percentages and number of individuals of the resolving groups identified with the use of a buffer zone.

Resolving groups identified using a single cut-off	Resolving groups identified using a buffer zone	Simulation without measurement error (ME), group % (*N*)	Simulation with ME, group % (*N*)
RC → no RD	RC → no RD	1.04 (15)	1.65 (24)
	RC → RC(bz)	0.77 (11)	0.82 (12)
	RC → RF(bz)	0.25 (4)	0.41 (6)
	RC → RF(bz) + RC(bz)	0.25 (4)	0.26 (4)
	RF(bz) + RC → no RD	0.27 (4)	0.42 (6)
	RF(bz) + RC → RC(bz)	0.21 (3)	0.24 (3)
	RF(bz) + RC → RF(bz)	0.35 (5)	0.36 (5)
	RF(bz) + RC → RF(bz) + RC(bz)	0.35 (5)	0.22 (3)
		3.49 (50)	**4.38 (63)**
RF → no RD	RF → no RD	0.39 (6)	0.93 (13)
	RF → RC(bz)	0.07 (1)	0.19 (3)
	RF → RF(bz)	0.97 (14)	1.06 (15)
	RF → RF(bz) + RC(bz)	0.24 (3)	0.26 (4)
	RF + RC(bz) → no RD	0.11 (2)	0.26 (4)
	RF + RC(bz) → RC(bz)	0.06 (1)	0.11 (2)
	RF + RC(bz) → RF(bz)	0.41 (6)	0.42 (6)
	RF + RC(bz) → RF(bz) + RC(bz)	0.22 (3)	0.19 (3)
		2.47 (36)	**3.42 (49)**
RF + RC → no RD	RF + RC → no RD	0.05 (1)	0.15 (2)
	RF + RC → RC(bz)	0.04 (1)	0.10 (1)
	RF + RC → RF(bz)	0.22 (3)	0.28 (4)
	RF + RC → RF(bz) + RC(bz)	0.21 (3)	0.20 (3)
		0.52 (8)	**0.73 (10)**

The number in parentheses shows the number of individuals based on the percentage for N = 1,432. The bold values show the total number of resolving cases.

few months of reading instruction in grade 1 (Lerkkanen et al., 2004), and by the time of grade 2 assessment [which was also the first assessment time point in the Catts et al. (2012) study], most are fluent readers and have a good command of reading comprehension skills (Lerkkanen et al., 2010). It is possible, then, that orthographic transparency could explain why there seem to be fewer cases of late-emerging RD among children learning to read in a context of transparent orthography as the differences in higher-level reading skills are visible already in grade 2. It is also likely that in a transparent orthography, it is possible to develop a resolving pathway more often despite early learning difficulties, as decoding task is cognitively less demanding. Additionally, Catts et al. (2012) followed the children until grade 8, whereas the present analysis extended only to grade 6. The longer gap between assessments may also increase the number of unstable RD cases. Finally, in our study, a somewhat stricter cut-off was used (10th percentile), while Catts et al. (2012) used the criterion of 1 standard deviation below the weighted sample mean (approximately 16th percentile).

Although the use of cut-offs is likely to lead to uncertainties in research findings because of measurement error (Francis et al., 2005; Branum-Martin et al., 2013), it is also a practical tool for the identification of children with RD. However, where and how we set the cut-off affects the identification of the RD groups and could possibly contribute to false impressions about the distinctness of the groups and about the risk of children for developing RD. Therefore, we need to find a way to use cut-offs without the evident problems accompanying them. One precaution against biased conclusions is the use of buffer zones around the single cut-offs (Shankweiler et al., 1999). The use of the buffer zone in the present study revealed a more complex picture than the one of the use of single cut-offs. For instance, the simulation without measurement error and without a buffer zone suggested that 4.38% of the children had late-emerging reading comprehension difficulties. However, when we use the buffer zone, we see that many of the children identified with late-emerging RD actually had borderline skills (lowest 10–25%) in reading fluency and/or reading comprehension already in grade 2. In other words, although many children passed the strict RD criterion, they nevertheless were still at the lower end of the skill distribution. Similarly, the simulation without the measurement error or buffer zone suggested that 2.48% of the children had resolving reading fluency difficulties, leading to the impression that these children were fluent readers in grade 6; in fact, most of them still scored in the borderline zone, just above the strict cut-off. The use of cut-off points is a practical tool for the identification of children with RD, but they can be problematic. Their use would be rational if we did not have a normal skill distribution, but this is not the case in reading achievement (Francis et al., 2005). Consequently, setting an arbitrary cut-off on the continuous distribution of reading achievement can lead to false or biased estimations of the prevalence of instability of the RD groups.

There are certain limitations in this study that need to be considered. First, we used only one measure for the assessment of reading comprehension in grades 2 and 6. Although we calculated the correction of attenuation in each grade in order to control measurement error, having more measures for the assessment of reading comprehension would have increased the strength of our model. Also, more measures and several time points would have allowed a more thorough and reliable assessment of reading comprehension.

In conclusion, this study shows that the use of measures with measurement error and the use of single cut-offs affect the longitudinal stability of RD identification across two time points. Comparing the prevalence of the groups arose from the use of single cut-off and those from the use of the buffer zone, it is evident that the use of single cut-off contributes to false impressions, such as how distinct the RD groups are. However, even after controlling for measurement error and using the buffer zone, our results suggest that RD are not stable over time for all children. Although many children manifest RD in the beginning of their school life and continue to have difficulties across grades, some children do not demonstrate difficulties until mid-primary school, and others may resolve their earlier difficulties by the end of primary school. Given that reading fluency and reading comprehension were not stable over time, the question arises about which additional factors affect the development of reading fluency and especially the development of reading comprehension, which was less stable. Further studies are needed to better understand the factors that could lead to late-emerging RD, either in reading fluency or in reading comprehension, as well as the factors that help children resolve their RD. Closer examination of the resolving cases could provide important information on the mechanisms that trigger protective factors. These insights could be used for the development of support systems and intervention programs, which will help children at risk for RD.

These results raise several clinical and practical implications. First, it seems that a change in the child's RD status can occur both because of the effect of measurement error and because of the instability of RD identification. Consequently, because of the presence of measurement error in every assessment tool, it is not sufficient to diagnose RD based on only one assessment. Although some children may pass the strict cut-off, the results of the buffer zone show that they may still be in jeopardy for RD. Second, the use of a buffer zone along with continuous follow-ups of children's reading development could facilitate more accurate identification of the children with RD. This is especially important in education systems in which access to remedial support or special needs interventions depends on an official diagnosis. Third, the accumulation of evidence for the instability of RD classification from this study and prior literature suggests a need for careful consideration of practices and permanency of diagnosing of RD. This is needed especially in cases where diagnostic practices deprive children with late-emerging RD of interventions or support and where individuals with resolving RD may continue to carry an inaccurate label or perception of one's skills.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the University of Jyväskylä. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MP and MT drafted the first version of the current manuscript. AT contributed to data analysis. M-KL and A-MP were responsible for data collection and commented on the manuscript. All authors contributed to the manuscript drafting, and read and approved the submitted version.

FUNDING

This research was supported by the Academy of Finland (Grant Numbers #213486, #268586, and #292466). MP was supported by the Alli Paasikiven Foundation. MT was supported by Academy of Finland (Grant Numbers #276239, #284439, and #313768).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02841/full#supplementary-material

REFERENCES

American Psychiatric Association, (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Association.

Aro, M. (2017). "Learning to read finnish," in *Reading Acquisition Across Languages and Writing Systems: An International Handbook* 1st Edn, eds L. T. W. Verhoeven, and C. A. Perfetti, (Cambridge: Cambridge University Press), 416–436. doi: 10.1017/9781316155752.017

Aro, M., and Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Appl. Psycholinguist.* 24, 621–635. doi: 10.1017/s0142716403000316

Björn, P. M., Aro, M. T., Koponen, T. K., Fuchs, L. S., and Fuchs, D. H. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learn. Disabil. Q.* 39, 58–66. doi: 10.1177/0731948715594787

Branum-Martin, L., Fletcher, J. M., and Stuebing, K. K. (2013). Classification and identification of reading and math disabilities: the special case of comorbidity. *J. Learn. Disabil.* 46, 490–499. doi: 10.1177/0022219412468767

Catts, H. W., Adlof, S. M., and Ellis Weismer, S. (2006). Language deficits in poor comprehenders: a case for the simple view of reading. *J. Speech Lang. Hear. Res.* 49, 278–293. doi: 10.1037/0708-5591.49.2.125

Catts, H. W., Compton, D., Tomblin, J. B., and Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *J. Educ. Psychol.* 104, 166–181. doi: 10.1037/a0025323

Catts, H. W., Hogan, T. P., and Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *J. Learn. Disabil.* 36, 151–164. doi: 10.1177/002221940303600208

Etmanskie, J. M., Partanen, M., and Siegel, L. S. (2016). A longitudinal examination of the persistence of late-emerging reading disabilities. *J. Learn. Disabil.* 49, 21–35. doi: 10.1177/0022219414522706

Florit, E., and Cain, K. (2011). The simple view of reading: is it valid for different types of alphabetic orthographies? *Educ. Psychol. Rev.* 23, 553–576. doi: 10.1007/s10648-011-9175-6

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., and Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD IQ and achievement scores are not sufficient. *J. Learn. Disabil.* 38, 98–108. doi: 10.1177/00222194050380020101

García, J. R., and Cain, K. (2014). Decoding and reading comprehension: a meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Rev. Educ. Res.* 84, 74–111. doi: 10.3102/0034654313499616

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remed. Spec. Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Holopainen, L. K., Kiuru, N. H., Mäkihonko, M. K., and Lerkkanen, M. K. (2018). The role of part-time special education supporting students with reading and spelling difficulties from grade 1 to grade 2 in Finland. *Eur. J. Spec. Needs Educ.* 33, 316–333. doi: 10.1080/08856257.2017.1312798

Hoover, W. A., and Gough, P. B. (1990). The simple view of reading. *Read. Writ.* 2, 127–160. doi: 10.1007/BF00401799

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struc. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Landerl, K., and Wimmer, H. (2008). Development of word reading fluency and spelling in a con-sistent orthography: an 8-year follow-up. *J. Educ. Psychol.* 100, 150. doi: 10.1037/0022-0663.100.1.150

Landerl, K., Wimmer, H., and Moser, E. (1997). *Salzburger Lese- und rechtschreibtest [Salzburg Reading and Spelling Test]*. Bern: Huber.

Leach, J. M., Scarborough, H. S., and Rescorla, L. (2003). Late-emerging reading disabilities. *J. Educ. Psychol.* 95, 211–224. doi: 10.1037/0022-0663.95.2.211

Lerkkanen, M.-K. (2007). "The beginning phases of reading literacy instruction in Finland," in *Finnish Reading Literacy. When Quality and Equity Meet*, eds P. Linnakylä, and I. Arffman, (Jyväskylä: University of Jyväskylä), 155–174. doi: 10.1108/s0735-004x(2010)0000023008

Lerkkanen, M.-K. (2018). "The influence of instruction on reading motivation in Finland," in *Reading achievement and Motivation in Boys and Girls: Field Studies and Methodological Approaches. Literacy Studies, 15*, eds P. O. Garcia, and P. B. Lind, (Cham: Springer), 65–78. doi: 10.1007/978-3-319-75948-7_4

Lerkkanen, M.-K. (2019). "Early language and literacy development in the Finnish context," in *The Sage Handbook of Developmental Psychology and Early Childhood Education*, eds D. Whitebread, V. Grau, K. Kumpulainen, M. M. McClelland, N. E. Perry, and D. Pino-Pasternak, (London: Sage), 403–417. doi: 10.4135/9781526470393.n23

Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, M., Siekkinen, M., and Nurmi, J.-E. (2006). *The First Steps Study (Ongoing)*. Finland: University of Jyväskylä.

Lerkkanen, M.-K., and Poikkeus, A.-M. (2009). *Lausetasoinen luetun ymmärtäminen ja sujuvuus: TOSREC-testin adaptoitu ja lyhennetty versio: Alkuportaat-tutkimuksen testimateriaalia [Sentence reading efficiency and comprehension: Adapted and shortened version of the TOSREC test: Test material of the First Steps study] [Unpublished test material]*. Finland: University of Jyväskylä.

Lerkkanen, M.-K., Poikkeus, A.-M., Ahonen, T., Siekkinen, M., Niemi, P., and Nurmi, J.-E. (2010). Luku- ja kirjoitustaidon sekä motivaation kehitys esi- ja alkuopetusvuosina [The development of reading and spelling skills from kindergarten to Grade 2]. *Kasvatus* 41, 116–128.

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., and Nurmi, J.-E. (2004). Predicting reading performance during the first and the second year of primary school. *Br. Educ. Res. J.* 30, 67–92. doi: 10.1080/0141192031000162997 4

Lindeman, J. (1998). *ALLU–Ala-Asteen Lukutesti [ALLU–Reading Test for Primary School]*. Turku: University of Turku.

Lindeman, J. (2000). *Ala-asteen lukutesti: Tekniset tiedot* ([2. p.].). *[ALLU Reading Test for Primary School: Technical Information]*. Turku: University of Turku.

Lipka, O., Lesaux, N., and Siegel, L. (2006). Retrospective analyses of the reading development of Grade 4 students with reading disabilities: risk status and profiles over 5 years. *J. Learn. Disabil.* 39, 364–378. doi: 10.1177/00222194060390040901

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Am. Stat. Assoc.* 83, 1198–1202. doi: 10.1080/01621459.1988.10478722

Lyytinen, H., Aro, M., Richardson, U., Erskine, J., Banff, A., Li, U. H., et al. (2015). "Reading skills, acquisition of: cultural, environmental, and developmental impediments," in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd Edn, ed. J. D. Wright, (Amsterdam: Elsevier), 5–11. doi: 10.1016/B978-0-08-097086-8.23111-5

Moll, K., Kunze, S., Neuhoff, N., Bruder, J., and Schulte-Körne, G. (2014). Specific learning disorder: prevalence and gender differences. *PLoS One* 9:e103537. doi: 10.1371/journal.pone.0103537

Nevala, J., and Lyytinen, H. (2000). *Sanaketjutesti [Differentiate Word Test]*. Jyväskylä: Niilo Mäki Instituutti.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition* 101, 385–413. doi: 10.1016/j.cognition.2006.04.008

Perfetti, C. A. (1985). *Reading Ability*. Oxford: Oxford University Press.

Pichler, C., and Wimmer, L. (2006). Das salzburger Lesescreening 2-9. Handreichnung für Lehrerinnen und Lehrer. Based on Mayringer, H., & Wimmer, H. (2003). Salzburger Lesescreening für die Klassenstufen 1-4 and Auer, M., Gruber, G., Mayringer, H., & Wimmer, H. (2005). *Salzburger Lesescreening für die Klassenstufen*. 5–8.

Santos, S., Cadime, I., Viana, F. L., and Ribeiro, I. (2019). Cross-lagged relations among linguistic skills in european portuguese: a longitudinal study. *Read. Res. Q.* doi: 10.1002/rrq.261

Schatschneider, C., Wagner, R. K., Hart, S. A., and Tighe, E. L. (2016). Using simulations to investigate the longitudinal stability of alternative schemes for classifying and identifying children with reading disabilities. *Sci. Stud. Read.* 20, 34–48. doi: 10.1080/10888438.2015.1107072

Seymour, P. H. K., Aro, M., and Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *Br. J. Psychol.* 94, 143–174. doi: 10.1348/000712603321661859

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., et al. (1999). Comprehension and decoding: patterns of association in children with reading difficulties. *Sci. Stud. Read.* 3, 69–94. doi: 10.1207/s1532799xssr0301_4

Snowling, M. J., and Hulme, C. (2012). Annual Research Review: the nature and classification of reading disorders–a commentary on proposals for DSM-5. *J. Child Psychol. Psychiatry* 53, 593–607. doi: 10.1111/j.1469-7610.2011.02495.x

Soodla, P., Lerkkanen, M. K., Niemi, P., Kikas, E., Silinskas, G., and Nurmi, J. E. (2015). Does early reading instruction promote the rate of acquisition? A comparison of two transparent orthographies. *Learn. Instruc.* 38, 14–23. doi: 10.1016/j.learninstruc.2015.02.002

Statistics Finland (2005). *Erityisopetusta Saavien Määrän Kasvu Jatkui [Number of Students Receiving Special Education has Decreased]*. Available at: http://www.stat.fi/til/erop/2004/erop_2004_2005-06-15_tie_001.html (accessed September 25, 2018).

Statistics Finland (2007). *Statistical Databases*. Available at: http://www.stat.fi/tup/tilastotietokannat/index_en.html (accessed September 25, 2018).

Torppa, M., Eklund, K., van Bergen, E., and Lyytinen, H. (2015). Late-emerging and resolving dyslexia: a follow-up study from age 3 to 14. *J. Abnorm. Child Psychol.* 43, 1389–1401. doi: 10.1007/s10802-015-0003-1

Torppa, M., Georgiou, G. K., Lerkkanen, M. K., Niemi, P., Poikkeus, A. M., and Nurmi, J. E. (2016). Examining the simple view of reading in a transparent orthography: a longitudinal study from kindergarten to grade 3. *Merrill-Palmer Q.* 62, 179–206. doi: 10.13110/merrpalmquar1982.62.2.0179

Torppa, M., Tolvanen, A., Poikkeus, A. M., Eklund, K., Lerkkanen, M. K., Leskinen, E., et al. (2007). Reading development subtypes and their early characteristics. *Ann. Dyslexia* 57, 3–32. doi: 10.1007/s11881-007-0003-0

Tunmer, W. E., and Chapman, J. W. (2012). The simple view of reading redux: vocabulary knowledge and the independent components hypothesis. *J. Learn. Disabil.* 45, 453–466. doi: 10.1177/0022219411432685

van Bergen, E., van der Leij, A., and de Jong, P. F. (2014). The intergenerational multiple deficit model and the case of dyslexia. *Front. Hum. Neurosci.* 8:346. doi: 10.3389/fnhum.2014.00346

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., and Pearson, N. A. (2009). *TOSREC: Test of Sentence Reading Efficiency and Comprehension*. Austin, TX: Pro-Ed.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Material

Appendix A
Mplus Input for Cross-Lagged Reading Model

```
TITLE: READING MODEL
DATA: FILE IS mplus.dat;
LISTWISE = ON;
VARIABLE: NAMES ARE childid cityid lukutaC8x difwC8c slsC8sum asiteksC8x
                                rtl2SC5 difwC5c rcsSC5 ly3C5c;
USEVARIABLES ARE lukutaC8x difwC8c slsC8sum asiteksC8x rtl2SC5 difwC5c rcsSC5
ly3C5c;
MISSING ARE ALL (-99);
ANALYSIS: ESTIMATOR=MLR;
                TYPE=GENERAL;
MODEL:   FlG2 BY RTL2SC5 DIFWC5C (1)
         RCSSC5 (2);
         FlG6 BY LUKUTAC8x DIFWC8C (1)
         SLSC8SUM (20);
         DIFWC5C WITH DIFWC8C;
         RTL2SC5 WITH LUKUTAC8x;
         RCSSC5 WITH SLSC8SUM;

!Kuder-Richardson reliability estimates for reading comprehension: ly3C5c: .80, asitetsC8x:
.74
! Variance ly3C5c=6.691
! Variance asitetsC8x=6.543
! Correction of attenuation (1-reliability)*variance
! For ly3C5c: (1-.80)*6.691=1.3382
! For asitetsC8x: (1-.74)*6.543=1.70118

         S1 BY LY3C5C@1; LY3C5C @1.3382;
         S1;
         S2 BY ASITEKSC8X@1; ASITEKSC8X @1.70118;
```

```
S2;
FlG6 ON FlG2 S1;
FlG2 WITH S1;
S2 ON S1 FlG2;
FlG6 WITH S2;

OUTPUT: SAMP STAND MOD(4) RES;
```

Appendix B

Syntax Simulation without Measurement Error and Single Cut-off

Note: The same syntax is used for the simulation without measurement error and the buffer zone but instead of one cut-off, we have two.

```
DATA LIST FREE/NR.
BEGIN DATA.
 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 (At this point we enter 200,000 free data. Because it
would take many pages to represent all 200,000, we have the first 20 as an example)
END DATA.

END DATA.
SET SEED=8949187.

*correlation is .549

COMPUTE COR1=NORMAL(1).
COMPUTE FLG2=SQRT(.549)*COR1+SQRT(1-.549)*NORMAL(1).
COMPUTE S1=SQRT(.549)*COR1+SQRT(1-.549)*NORMAL(1).
EXECUTE.

*residual correlation is .131
*residual variances are .250 for fluency and .593 for comprehension
*RESIDUALS

COMPUTE COR2=NORMAL(1).
COMPUTE RES_FL6=SQRT(.131)*COR2+SQRT(1-.131)*NORMAL(1).
COMPUTE RES_S2=SQRT(.131)*COR2+SQRT(1-.131)*NORMAL(1).
CORRELATIONS
    /VARIABLES=RES_FL6 RES_S2
    /PRINT=TWOTAIL NOSIG
    /MISSING=PAIRWISE.

COMPUTE FLG6=0.782*FLG2+0.139*S1+SQRT(0.250)*RES_FL6.
COMPUTE S2=0.568*S1+0.114*FLG2+SQRT(0.593)*RES_S2.
EXECUTE.

DESCRIPTIVES VARIABLES=FLG2 S1 FLG6 S2
  /STATISTICS=MEAN STDDEV MIN MAX.

SET DECIMAL=DOT.

WRITE OUTFILE='<DESTINATION TO SAVE NEW DATASET>' ENCODING='ASCII'/
    FLG2 FLG6 S1 S2 (4F12.3).
EXECUTE.

SORT CASES BY FLG2.
EXECUTE.
IF ($CASENUM LE 20000) FLG2_DIK=1.
```

```
IF ($CASENUM GT 20000) FLG2_DIK=0.
EXECUTE.
SORT CASES BY FLG6.
EXECUTE.
IF ($CASENUM LE 20000) FLG6_DIK=1.
IF ($CASENUM GT 20000) FLG6_DIK=0.
EXECUTE.
SORT CASES BY S1.
EXECUTE.
IF ($CASENUM LE 20000) S1_DIK=1.
IF ($CASENUM GT 20000) S1_DIK=0.
EXECUTE.
SORT CASES BY S2.
EXECUTE.
IF ($CASENUM LE 20000) S2_DIK=1.
IF ($CASENUM GT 20000) S2_DIK=0.
EXECUTE.

GENLOG FLG2_DIK S1_DIK FLG6_DIK S2_DIK
  /MODEL=POISSON
  /PRINT=FREQ RESID ADJRESID ZRESID DEV
  /PLOT=RESID(ADJRESID) NORMPROB(ADJRESID)
  /CRITERIA=CIN(95) ITERATE(20) CONVERGE(0.001) DELTA(.5)
  /DESIGN FLG2_DIK S1_DIK FLG6_DIK S2_DIK.
```

Appendix C

Syntax Simulation with Measurement Error and Single Cut-off

Note: The same syntax is used for the simulation with measurement error and the buffer zone but instead of one cut-off, we have two.

```
DATA LIST FREE/NR.
BEGIN DATA.
 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 (At this point we enter 200,000 free data. Because it would take many pages to represent all 200,000, we have the first 20 as an example)
END DATA.

SET SEED=37402897.

*correlation is .549

COMPUTE COR1=NORMAL(1).
COMPUTE FLG2=SQRT(.549)*COR1+SQRT(1-.549)*NORMAL(1).
COMPUTE S1=SQRT(.549)*COR1+SQRT(1-.549)*NORMAL(1).
EXECUTE.

*residual correlation is .131
* residual variances are .250 for fluency and .593 for comprehension
* RESIDUALS

COMPUTE COR2=NORMAL(1).
COMPUTE RES_FL6=SQRT(.131)*COR2+SQRT(1-.131)*NORMAL(1).
COMPUTE RES_S2=SQRT(.131)*COR2+SQRT(1-.131)*NORMAL(1).
CORRELATIONS
   /VARIABLES=RES_FL6 RES_S2
   /PRINT=TWOTAIL NOSIG
   /MISSING=PAIRWISE.

COMPUTE FLG6=.782*FLG2+.139*S1+SQRT(.250)*RES_FL6.
COMPUTE S2=.568*S1+.114*FLG2+SQRT(.593)*RES_S2.
EXECUTE.

DESCRIPTIVES VARIABLES=FLG2 S1 FLG6 S2
   /STATISTICS=MEAN STDDEV MIN MAX.

COMPUTE REC1=NORMAL(1).
COMPUTE REC2=NORMAL(1).
COMPUTE REC3=NORMAL(1).
EXECUTE.
COMPUTE RES1=SQRT(.073)*REC1+SQRT(1-.073)*NORMAL(1).
COMPUTE RES2=SQRT(.262)*REC2+SQRT(1-.262)*NORMAL(1).
COMPUTE RES3=SQRT(.264)*REC3+SQRT(1-.264)*NORMAL(1).

COMPUTE F21=.887*FLG2+SQRT(.213)*RES1.
```

```
COMPUTE F22=.597*FLG2+SQRT(.644)*RES2.
COMPUTE F23=.746*FLG2+SQRT(.444)*RES3.
COMPUTE S11=.894*S1+SQRT(.200)*NORMAL(1).
EXECUTE.
COMPUTE RES4=SQRT(.073)*REC1+SQRT(1-.073)*NORMAL(1).
COMPUTE RES5=SQRT(.262)*REC2+SQRT(1-.262)*NORMAL(1).
COMPUTE RES6=SQRT(.264)*REC3+SQRT(1-.264)*NORMAL(1).

COMPUTE F61=.852*FLG6+SQRT(.274)*RES4.
COMPUTE F62=.670*FLG6+SQRT(.551)*RES5.
COMPUTE F63=.736*FLG6+SQRT(.458)*RES6.
COMPUTE S21=.860*S2+SQRT(.260)*NORMAL(1).
EXECUTE.

COMPUTE RF2SUM=F21+F22+F23.
COMPUTE RF6SUM=F61+F62+F63.
COMPUTE LY3C5C=S11.
COMPUTE ASITEKSC8X=S21.
EXECUTE.

SET DECIMAL=DOT.

WRITE OUTFILE='<DESTINATION TO SAVE NEW DATASET>' ENCODING='ASCII'/
   F21 F22 F23 S11 F61 F62 F63 S21 (8F12.3).
EXECUTE.

SORT CASES BY RF2SUM.
EXECUTE.
IF ($CASENUM LE 20000) RF2_SDIK=1.
IF ($CASENUM GT 20000) RF2_SDIK=0.
EXECUTE.
SORT CASES BY RF6SUM.
EXECUTE.
IF ($CASENUM LE 20000) RF6_SDIK=1.
IF ($CASENUM GT 20000) RF6_SDIK=0.
EXECUTE.
SORT CASES BY LY3C5C.
EXECUTE.
IF ($CASENUM LE 20000) CO2_SDIK=1.
IF ($CASENUM GT 20000) CO2_SDIK=0.
EXECUTE.
SORT CASES BY ASITEKSC8X.
EXECUTE.
IF ($CASENUM LE 20000) CO6_SDIK=1.
IF ($CASENUM GT 20000) CO6_SDIK=0.
EXECUTE.

GENLOG RF2_SDIK CO2_SDIK RF6_SDIK CO6_SDIK
   /MODEL=POISSON
   /PRINT=FREQ RESID ADJRESID ZRESID DEV
```

```
/PLOT=RESID(ADJRESID) NORMPROB(ADJRESID)
/CRITERIA=CIN(95) ITERATE(20) CONVERGE(0.001) DELTA(.5)
/DESIGN RF2_SDIK CO2_SDIK RF6_SDIK CO6_SDIK.
```

III

DEVELOPMENTAL PROFILES OF READING FLUENCY AND READING COMPREHENSION FROM GRADES 1 TO 9 AND THEIR EARLY IDENTIFICATION

by

Maria Psyridou*, Asko Tolvanen, Peter de Jong, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus, & Minna Torppa

Submitted manuscript

Request a copy from the author.