Riitta Virinkoski

Teachers' Practices and Competence to Assess Students' Academic Skills in Primary School





Riitta Virinkoski

Teachers' Practices and Competence to Assess Students' Academic Skills in Primary School

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella julkisesti tarkastettavaksi maaliskuun 11. päivänä 2022 kello 12.

Academic dissertation to be publicly discussed, by permission of the Faculty of Education and Psychology of the University of Jyväskylä, on March 11, 2022 at 12 o'clock.



JYVÄSKYLÄ 2022

Editors Miika Marttunen Department of Teacher Education, University of Jyväskylä Timo Hautala Open Science Centre, University of Jyväskylä





Cover picture by Juulia Lempinen

Copyright © 2022, by University of Jyväskylä

ISBN 978-951-39-9032-9 (PDF) URN:ISBN:978-951-39-9032-9 ISSN 2489-9003

Permanent link to the online version of this publication: http://urn.fi/URN:ISBN:978-951-39-9032-9

ABSTRACT

Virinkoski, Riitta
Teachers' Practices and Competence to Assess Students' Academic Skills in Primary School
Jyväskylä: University of Jyväskylä, 2022, 81 p.
(JYU Dissertations
ISSN 2489-9003; 495)
ISBN 978-951-39-9032-9 (PDF)

The aim of the present thesis was to introduce the findings concerning primary school teachers' assessment practices of students' academic skills, the accuracy of the assessments, and the connection between the length of part-time special education and the development of students' academic skills. First, it examined class teachers and special education needs (SEN) teachers' assessment practices (tests, qualitative assessments, and curriculum-based measures), and their ability to assess reading in Grades 1 and 6 (SEN teachers only), and second, it explored the relation between the development of students' academic skills across Grades 1 to 4 and the reasons for and the length of part-time special education. The data for the sub-studies were drawn from two Finnish longitudinal studies: Jyväskylä Longitudinal Study of Dyslexia (JLD) and the First Steps study. The samples in the sub-studies were: 103 teachers and 113 students in Grade 1 (study I), 29 teachers and 55 students in Grade 6 (study II), and 35 teachers and 130 students in Grades 1 to 4 (study III). The data comprised teacher questionnaires, student ratings, and students' test scores of literacy and math skills. The data were analyzed using statistical methods. The results showed that in Grade 1, most class teachers used one assessment practice, while SEN teachers usually combined two practices. In Grade 6, SEN teachers preferred multiple practices, but qualitative practices were considered the most important. Associations between the teacher ratings and test scores were mostly moderate, and typically performing students were identified better than struggling students, except in study I where SEN teachers identified all at-risk students in Grade 1. The identification of typically performing students in reading comprehension was the most challenging for the SEN teachers in Grade 6. Finally, the development of literacy and math skills among the students in the long-lasting part-time special education support group across Grades 1 to 4 lagged behind the other two groups of students (Study III). The findings indicate that to gain a comprehensive view of the students' skills, using various assessment practices is recommended. And after identifying the risk or learning difficulty, more attention should be paid to monitoring skill development.

Keywords: teacher assessment, accuracy, part-time special education, learning difficulties, academic skill development, primary school students

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Virinkoski, Riitta

Opettajien arviointimenetelmät ja arviointiosaaminen oppilaiden akateemisten taitojen arvioinnissa alakoulussa

Jyväskylä: University of Jyväskylä, 2022, 81 s.

(JYU Dissertations ISSN 2489-9003; 495)

ISBN 978-951-39-9032-9 (PDF)

Tässä väitöskirjassa käsiteltiin alakoulun opettajien akateemisten taitojen arviointimenetelmiä, arviointien tarkkuutta sekä osa-aikaisen erityisopetuksen keston vaikutusta oppilaiden taitojen kehittymiseen. Tavoitteena oli tutkia luokanja erityisopettajien arviointimenetelmiä, eli laadullista arviointia, testiarviointia ja opetussuunnitelmapohjaista arviointia, ja kykyä arvioida oppilaiden lukutaitoa ensimmäisellä ja kuudennella (vain erityisopettajat) luokalla sekä 1-4-luokkien oppilaiden akateemisten taitojen kehityksen ja osa-aikaisen erityisopetuksen syiden ja keston välistä suhdetta. Osatutkimusten aineistot olivat peräisin kahdesta suomalaisesta pitkittäistutkimuksesta: Lapsen Kielen Kehitys (LKK) ja Alkuportaat. Otokset osatutkimuksissa olivat 103 opettajaa ja 113 oppilasta luokalla 1 (tutkimus I), 29 opettajaa ja 55 oppilasta luokalla 6 (tutkimus II) sekä 35 opettajaa ja 130 oppilasta luokka-asteilla 1-4 (tutkimus III). Aineisto koostui opettajien kyselylomakkeista, opettajien tekemistä oppilasarvioista ja oppilaiden lukutaidon ja matematiikan testituloksista. Aineisto analysoitiin tilastollisia menetelmiä käyttäen. Tulokset osoittivat, että 1. luokalla luokanopettajat käyttivät vain yhtä arviointimenetelmää, erityisopettajat puolestaan yleensä kahta menetelmää. Erityisopettajat (tutkimus II) suosivat useita arviointitapoja, mutta laadullisia menetelmiä pidettiin tärkeimpänä. Opettajien arvioiden ja oppilaiden testitulosten väliset yhteydet olivat enimmäkseen kohtalaisia, ja ikätasolle tyypillisesti suoriutuvat oppilaat tunnistettiin paremmin kuin lukemisessaan vaikeuksia kohtaavat oppilaat, paitsi että Alkuportaiden erityisopettajat tunnistivat kaikki 1. luokan oppilaat, joilla oli riski lukivaikeuksiin. Kuudennella luokalla erityisopettajien oli vaikeinta tunnistaa oppilaita, jotka suoriutuivat ikätasolle tyypillisesti luetun ymmärtämisessä. Lukemisen ja matematiikan taitojen kehitys luokilla 1-4 oli hitaampaa pitkäkestoisen erityisopetuksen tukiryhmässä verrattuna kahteen muuhun ryhmään (tutkimus III). Tulokset osoittivat, että kokonaisvaltaisen näkemyksen saamiseksi oppilaiden akateemisista taidoista erilaisia arviointimenetelmiä tulee käyttää monipuolisesti. Lisäksi riskin tai oppimisvaikeuden tunnistamisen jälkeen olisi kiinnitettävä enemmän huomiota taitojen kehityksen seurantaan.

Asiasanat: opettajien tekemä arviointi, tarkkuus, osa-aikainen erityisopetus, oppimisvaikeudet, akateemisten taitojen kehitys, alakoulun oppilaat

Author Riitta Virinkoski

Open University

P.O. Box 35

FI-40014, University of Jyväskylä, Finland

riitta.e.virinkoski@jyu.fi

http://orcid.org/0000-0002-7270-8230

Supervisors Professor Mikko Aro

Department of Education

University of Jyväskylä, Finland

Professor Marja-Kristiina Lerkkanen Department of Teacher Education University of Jyväskylä, Finland

Reviewers Professor Marita Mäkinen

Faculty of Education and Culture University of Tampere, Finland

Professor Krista Uibu Institute of Education

University of Tartu, Estonia

Opponent Professor Marita Mäkinen

Faculty of Education and Culture University of Tampere, Finland

ACKNOWLEDGEMENTS

I would never have thought that I could have proceeded this far in my studies if there had not been a certain coincidence. This whole idea of my doctoral studies hit me like a "bolt from the blue" during a discussion with Professor Mikko Aro after having recorded a video of reading difficulties for the Open University. The timing was ideal for me then, as I was "between jobs" as a university teacher and did not yet have a permanent teacher position. Many times, during these years, I wondered "what on earth am I doing" and how everything is finally going to turn out.

I have had the wonderful opportunity to work with the most respected and experienced supervisors, and for that, I express my deepest gratitude to Professor Mikko Aro, Professor Marja-Kristiina Lerkkanen, and also my co-authors, Dr. Kenneth Eklund and Professor Leena Holopainen. I thank both my supervisors for being patient with me and believing in me, even at times when I did not. I cannot thank you enough for your insightful comments and support along the way. I am also extremely grateful to Kenneth for his interest in and contribution to the research data and especially the analyses, and for giving me valuable comments on all three manuscripts. I was honored to have Professor Marita Mäkinen and Professor Krista Uibu as evaluators and pre-examiners of my thesis. Your encouraging comments and well-justified views helped me to add the finishing touches to this thesis. I am most grateful to Professor Mäkinen, who agreed to act as an opponent in the public defense. I also thank Docent Matti Kuorelahti for acting as my "follow-up group member" during this process.

In my thesis, I have been able to use the data sets from two highly valued longitudinal studies (Jyväskylä Longitudinal Study of Dyslexia, and First Steps study), and I express my deepest gratitude to all people involved with those studies. I also had the opportunity to take part in gathering the study data in 2013 from the special education teachers in the First Steps study. It has been an immense honor to be part of such an important research project. Additionally, the assistance from Jenni Ruotsalainen with the First Steps study data sets along the way was invaluable to me as it enabled me to effectively focus on each of the articles.

Although I have almost all of this time been working full-time as a university teacher, I have also had the opportunity to get funding for a few 100% research periods from the First Steps follow-up study funded by the Academy of Finland, the Ellen and Artturi Nyyssönen Foundation, the LUKILOKI teacher professional development program funded by the National Board of Education, and from the Department of Education. This support has made it possible for me to combine doctoral studies and work and to successfully complete my thesis, for which I am most grateful.

Finishing this project demanded a great deal of work and determination, not to mention time. But I am not a person who leaves things unfinished once I have decided to start something. My feelings have varied from frustrated and desperate to hopeful, sometimes even successful. However, working with my

doctoral thesis has been worth every experience; I can see it clearly now. My dear co-workers at the Open University of the University of Jyväskylä, especially all the members of the special education team (Liisa, Sanna, Tarja, Tiina, Erika, Katja L., Katja P., and Kristiina), have supported me through all these years. Their flexibility concerning the practical arrangements during my absence from our teacher team has been amazing and heart-warming. Additionally, the director of the Open University, Jukka Lerkkanen's supportive and encouraging attitude has inspired me to complete this project, so I thank him. I have also had the opportunity, in the form of funding from the Open University, to take part in some international conferences that supported my doctoral studies, for which I am most grateful. Additionally, the discussions and presentations in the doctoral seminar group over the years have broadened my own ideas and given me the peer support to carry on with this project.

Last, but not least, I warmly thank all my family members and friends for their precious support and understanding along the way. Particularly Maarit, Nina, Annamaija, and Titta, being able to discuss my doctoral studies with them, among other things, has been most rewarding. I dedicate this thesis to my three wonderful children: Emilia, Juulia, and Rasmus as well as my three grandchildren, Minea, Sofia, and Enni.

Jyväskylä, February 2022 Riitta Virinkoski

LIST OF PUBLICATIONS

This doctoral thesis is based on the following publications, which are referred to as sub-studies in this doctoral study. The research articles are reprinted with the permission of the publishers. Copies of the articles are appended to the thesis.

- Article I Virinkoski, R., Lerkkanen, M.-K., Holopainen, L., Eklund, K., & Aro, M. (2018). Teachers' ability to identify children at early risk for reading difficulties in Grade 1. *Early Childhood Education Journal*, 46(5), 497–509. https://doi.org/10.1007/s10643-017-0883-5
- Article II Virinkoski, R., Lerkkanen, M.-K., Eklund, K., & Aro, M. (2020). Special education teachers' identification of students' reading difficulties in Grade 6. Scandinavian Journal of Educational Research, 66(1), 59–72. https://doi.org/10.1080/00313831.2020.1833241
- Article III Virinkoski, R., Eklund, K., Lerkkanen, M.-K., Holopainen, L. & Aro, M. (2020). Development of reading and arithmetic skills across Grades 1 to 4 in two groups of children receiving part-time special education. *Learning and Individual Differences*, 85, 101956. https://doi.org/10.1016/j.lindif.2020.101956

The author of this thesis is the first author of all three articles. She was responsible for the sub-studies design, searching and reviewing the literature, and writing the manuscripts. She was also responsible for selecting the relevant research data as well as collecting SEN teacher data from Grade 6 in the First Steps study. She carried out the statistical analyses in collaboration with a statistical expert who is also one of the co-authors. The co-writers had advisory roles in designing the study, commenting on the manuscripts, and interpreting the results.

FIGURES

FIGURE 1	Specific research aims and research questions of the thesis	30
FIGURE 2	Samples in sub-study I.	34
FIGURE 3	Samples in sub-study II	34
FIGURE 4	Samples in sub-study III	
TABLES		
TABLE 1	Measures in sub-studies I-III	38
TABLE 2	Overview of the themes, variables, and statistical methods in	40

CONTENTS

ABSTRACT
TIIVISTELMÄ (ABSTRACT IN FINNISH)
ACKNOWLEDGEMENTS
LIST OF PUBLICATIONS
FIGURES AND TABLES
CONTENTS

1	INT	RODUCTION	13
2	THI	EORETICAL BACKGROUND	17
2.	2.1	Teachers as assessors	17
	2.2	Assessment and academic skills' development	19
		2.2.1 Forms of assessment	19
		2.2.2 Subject of assessment	20
	2.3	Teachers' assessment of literacy and math	23
	2.4	Accuracy of teachers' assessment	
	2.5	The aims of the thesis	28
3	ME	THOD	31
	3.1	Participants and data	31
	3.2	Measures	36
	3.3	Analytical strategies	39
4	OVI	ERVIEW OF THE ORIGINAL STUDIES	41
	4.1	Sub-study I: Teachers' ability to identify children at early risk	
		for reading difficulties in Grade 1	41
	4.2	Sub-study II: Special education teachers' identification of	
		students' reading difficulties in Grade 6	42
	4.3	Sub-study III: Development of reading and arithmetic skills	
		across Grades 1 to 4 in two groups of children receiving	
		part-time special education	43
5	GEN	NERAL DISCUSSION	
	5.1	Teachers' assessment practices to identify reading difficulty	45
	5.2	Accuracy of teachers' assessments	
	5.3	Learning support and development of academic skills	49
	5.4	Ethical considerations	52
	5.5	Practical implications	52
	5.6	Limitations and future directions	54
6	COI	NCLUDING REMARKS	57
VΗ	TEEN	VETO	57

REFERENCES	63
ORIGINAL PAPERS	82

1 INTRODUCTION

Learning to read is one of the most important life skills regarding one's academic achievement across compulsory education, which also affects an individual's career choices and occupational possibilities in the future (Eloranta et al., 2019). Teachers' assessments of students' literacy and math skills and of the possible risk for learning difficulties (LDs) are important, especially at the beginning of primary school. For example, depending on the criteria, 5%–18% of students experience reading difficulties (RDs) at some stage of primary school, and for some students, these difficulties can be rather persistent (Fedora, 2015; Holopainen, Kiuru et al., 2017; Shaywitz & Shaywitz, 2005). It has been indicated that when students start to lag behind their peers in their development, they generally tend to stay behind (Fedora, 2015; Snow et al., 1998; Torppa et al., 2015). Therefore, early recognition of the risk for RD is necessary to avoid prolonged or more serious LDs later.

Prior studies have indicated that teachers' assessment accuracy compared with the scores of standardized achievement tests is systematically lower concerning struggling students than typically performing students (Meissel et al., 2017) and that teachers are more accurate in judging typically performing students (Coladarci, 1986; Soodla & Kikas, 2010). However, a recent review of studies dealing with the accuracy of teacher judgments (Urhahne & Wijnia, 2021) and a replication of a former meta-analysis (Hoge & Coladarci, 1989; also see Kaufmann, 2020) have provided updated knowledge of teachers' assessment accuracy and the moderators connected with it.

The Finnish language is orthographically one of the most consistent languages displaying regular and symmetrical grapheme-phoneme correspondences, which facilitates rapid mastery of decoding accuracy (Aro, 2017). Compulsory pre-primary education in Finland starts at age six, and at that point, the children become acquainted with graphemes and phonemes (Lerkkanen et al., 2004) as well as mathematical concepts and numbers (Aunola et al., 2004). However, pre-primary education does not involve formal instruction in literacy or mathematics. Only when the children enter the first grade at age seven does formal

instruction of reading, spelling, and arithmetic begin. In Grade 1, reading and spelling are taught simultaneously, which is based on the scientific knowledge related to transparent orthographies that development in one skill complements the development of the other (Lerkkanen, 2007).

The objectives of the Finnish educational policy are to guarantee equal opportunities for every child to be educated and to be able to enroll in a local school (Government of Finland, 2010). The Finnish National Core Curriculum (Finnish National Board of Education [FNBE], 2004, 2014) describes the mission, values, and structure of basic education, and it has three main functions: being an administrative, an intellectual, and a pedagogical document (Niemi et al., 2012; Sullanmaa et al., 2019; Vitikka et al., 2012). Of these three, the core curriculum as a pedagogical document is a tool for teachers providing pedagogical advice and support, and setting guidelines for teaching and learning, on which the teachers can build their own pedagogical practices (Vitikka et al., 2012). The core curriculum is also the framework for the local curricula, and municipalities, schools, and teachers have broad autonomy in implementing the core curriculum and organizing education (Halinen & Järvinen, 2008; Kyttälä et al., 2021).

Finnish teachers are considered trustworthy professionals and are given the responsibility for their students' academic development (Lerkkanen et al., 2013; Penttinen et al., 2020). All teachers in Finland have a master's degree, but training differs in gaining competence in identifying at-risk students (Takala & Ahl, 2014). Finnish special education needs (SEN) teachers are trained to identify students' difficulties in reading, writing, language development, and mathematical and behavioral issues, and to tailor the support practices for each student, based on the student evaluations (Björn et al., 2016; Takala et al., 2009). No national assessments are used in Finland, and at the school level, the eligibility for support is assessed multi-professionally, generally on the grounds of the class teachers, SEN teachers, and the parents' observations and views (Björn et al., 2016; Jahnukainen & Itkonen, 2016; 2018; Takala et al., 2009). This differs from other countries. For example, in the United States, identification of at-risk students depends strongly on frequent, systematic psychometric testing and screening pertaining to special education programs (Fuchs et al., 2012).

Until 2010, in Finland, special education comprised two different forms of special education: part-time special education and full-time special education, the latter referring to education for students accepted and transferred to special education (Pulkkinen & Jahnukainen, 2016; Pulkkinen et al., 2020). Students with mild LDs were provided with part-time special education, while students needing more regular and intensive support had to be transferred to special education (full-time special education), and an official decision for special education and an individual education plan (IEP) were required (Pulkkinen & Jahnukainen, 2016). Legislation and regulations of provision for learning and schooling support within the Finnish education system were renewed in 2010. The changes were officially adopted in August 2011, as a consequence of the current Basic Education Act (Finnish Government, 2010) with the National Core Curriculum (FNBE, 2010) coming into effect, and the mandatory, tiered three-level framework for

15

support was introduced (Björn et al., 2016). The data used in this thesis were, for the most part, gathered during the earlier legislation.

The current Finnish tiered support model consists of three tiers: tier 1 (general support), tier 2 (intensified support), and tier 3 (special support) (Jahnukainen & Itkonen, 2018). All students attending basic education are entitled to appropriate and timely support for their learning whenever they need it, and all forms of support (i.e., differentiation, remedial teaching, part-time special education, an assistant) are feasible at any tier of support, except full-time special education in tier 3 (FNBE, 2014; Jahnukainen & Itkonen, 2018). During primary school, class teachers are usually the first to provide general support to students with difficulties in reading and mathematics. If that support is deemed to be insufficient, support is provided by the SEN teachers, usually in a resource room either individually or in small groups (Holopainen, Kiuru, et al., 2017).

The school system in Finland differs in several ways from that of many other countries, one distinctive characteristic being that the learning and schooling support is provided within the mainstream schools, generally free of charge. To receive special education services, pedagogical grounds are considered adequate, and the student does not need a formal diagnosis, nor do formal criteria for support exist (Björn et al., 2016; Jahnukainen & Itkonen, 2018). Part-time special education as a typical form of support is rather extensive when compared internationally: in 2019, the share of students receiving this kind of support at some point in the school year from the SEN teachers was 22% across comprehensive schools (Official Statistics of Finland [OSF], 2019). Nevertheless, the conditions of the share of students receiving part-time special education and the ways that support is organized vary among the municipalities (Pulkkinen et al., 2020). The most common reasons for part-time special education in primary school are difficulties in reading and spelling, expressive language disorders, and arithmetic difficulties (Holopainen, Kiuru, et al., 2017; OSF, 2019; Savolainen et al., 2018). Part-time special education is provided within every level of the three-tiered framework for support, and there are no specific guidelines for assessment procedures, intensity, duration, or content of support (Björn et al., 2016; The Basics of the curriculum in Finnish basic education [Perusopetuksen opetussuunnitelman perusteet] (FNBE, 2014). Even though part-time special education was originally intended as a temporary, low-threshold form of learning support for any student, it can be provided as long as the support is considered necessary (Björn et al., 2016; Holopainen, Kiuru et al., 2017).

In the Finnish school system, class teachers and SEN teachers play a focal role in the assessment of students' academic skills from very early on and in providing support across different phases of the school path. Despite numerous studies addressing the identification of students' LDs using various measures, such as standardized tests and other direct assessment tools, not to mention modern computer-based applications (Auphan et al., 2019; Gaab et al., 2019; Hautala et al., 2020; Shapiro et al., 2017), the examination of the role of the teachers' ratings in assessment of academic skills' assessment still needs more attention. There is a knowledge gap regarding teachers' (class and subject teachers, SEN teachers)

assessment practices: what they are, how functional they are, and how accurate they are in identifying the students' LDs and in helping the planning of individual support. Provision of teachers' support resources is limited, which also emphasizes the accurate identification of typically performing students who do not need teachers' support in learning, enabling targeting the support to students who need it the most.

Currently, the knowledge about how the students' academic skills (reading and math) and learning difficulties are being assessed by the teachers in primary school, and what is the outcome of the assessments, is rather limited. There are also quite few studies (e.g., Holopainen, Kiuru et al., 2017) examining the effect of part-time special education on the development of students' academic skills. The aim of this thesis was to gain new knowledge of how class teachers and SEN teachers assess school beginners' pre-reading skills at school entry, and students' levels of reading fluency and reading comprehension at the end of primary school, in Grade 6. The focus was, in addition to exploring the teachers' assessment practices, on examining the accuracy of the teachers' ratings as compared with the students' test scores. The thesis also aimed to determine the accuracy of the identification of struggling and typically (according to age-level) performing students, and how well the teachers were able to identify the students developing typically and the students at risk for RDs, or who already manifested RD. Also, SEN teachers' reported reasons for part-time special education in Grade 1, together with their reports of students' overlapping difficulties (e.g., attention difficulty together with RD), were investigated. Additionally, how the groups with different lengths of part-time special education (1-2 years, 3-4 years) differed from each other and their peers in the control group in terms of reading fluency, reading comprehension, and arithmetic fluency in Grades 1 and 4 were examined. Also, the relation between the development of students' aforementioned academic skills across Grades 1 to 4 and the reasons as well as the length of parttime special education periods were studied.

2 THEORETICAL BACKGROUND

2.1 Teachers as assessors

Prior literature presents various concepts associated with teachers' assessment of students' skills. Assessment literacy (Stiggins, 1991, 1995, 2002) is a broad, traditional concept encompassing both teachers' assessment knowledge and skills associated with teacher practice. It refers to teachers' professional competency in educational assessment (DeLuca et al., 2016; Pastore & Andrade, 2019; Xu & Brown, 2016), basic understanding of educational assessment, and the related skills for using their knowledge of various assessment measures concerning student achievement (Herppich et al., 2018; Xu & Brown, 2016). However, according to prior studies, teachers often lack sufficient background or training in assessment, even though they are involved in assessment-related decision-making situations (Popham, 2009; Xu & Brown, 2016). Assessment literacy involves the ability to construct reliable assessments and then administer and score the assessments in order to facilitate valid instructional decisions connected with provincial or state standards (DeLuca et al., 2016). Assessment literacy helps teachers in their work via perceiving, analyzing, and using data on student performance, and poor assessment literacy (illiteracy) can lead to inaccuracy in the assessments, and thus, the purpose of the assessments cannot be fulfilled (Bayat & Rezaei, 2015). Also, assessment literacy supports teachers in responding to students' learning needs by providing more effective teaching based on the information about student learning (Pastore & Andrade, 2019).

Assessment competency or diagnostic competence is also regarded as a foundation of teachers' expertise (Herppich et al., 2018; Ohle & McElvany, 2015; Schrader, 2013). According to Herppich et al. (2018), a generic term for the accuracy of teacher judgments, diagnostic competence, includes teachers' knowledge about judgment processes, individual abilities and difficulties of the students,

and the necessary methodological and procedural knowledge needed for assessing student performance. A conceptual model of teachers' assessment competence by Herppich et al. (2018) contains assessment products (= judgments), as outcomes of assessment processes, and assessment practices as well as teachers' knowledge, beliefs and motivations, and assumes that competences are learnable (see also Glogger-Frey et al., 2018). In the model, teachers' professional knowledge is linked to real-life performance in pedagogical situations (Herppich et al., 2018). This model presumes that certain overarching characteristics combine high assessment competence situations: starting prior to educational decisions or actions, assessments are subordinate to educational decisions, including information for the decision, assessment practices referring to the students, and addressing the characteristics relevant to learning (Glogger-Frey, 2018; Herppich et al., 2018).

Prior studies have indicated that teachers may have varying assessment competence, and long working experience with a certain class or more contacts with each student do not necessarily increase teachers' assessment accuracy (Herppich et al., 2018; Urhahne & Wijnia, 2021). It has been suggested that in addition to assessment knowledge and skills, teachers need to have positive beliefs about the assessment data as well as the capacity to use data in their daily instructional decisions (Datnow & Hubbard, 2016; McKevett & Kiss, 2019). Teachers' conceptions of assessment build on their intuitive understanding of assessment, including beliefs, thoughts, prior information, and feelings regarding assessment (Kyttälä et al., 2021; see also Xu & Brown, 2016). It has been shown that pre-service teachers' individual conceptions of assessment, prior work experience, and theoretical knowledge affect not only the ways that they assess the students but also the interpretations and conclusions they make, and some preservice teachers consider assessment to have a detrimental influence on student learning (Kyttälä et al., 2021).

Making educational decisions and planning instruction and support require data collection and monitoring of the students' learning progress (Herppich et al., 2018; Südkamp et al., 2018). A recent study by Al Otaiba et al. (2019) revealed that teachers' knowledge related to gathered assessment information was lacking, and this complicated making instructional decisions for the students at risk for RD. It has been found that there is considerable variation in teachers' data collection practices, ranging from rational, deliberate searching for data to intuitive, recognition-based practices (Vanlommel et al., 2018). Teachers can collect data using intuitive processes, which highlight the value of the personal knowledge of experts, and through spontaneous recognition of cues (Klein, 2008; Vanlommel & Schildkamp, 2019; Vanlommel et al., 2018), or in a nonsystematic manner, for example, using on-the-fly assessments or observing students (Heritage, 2007; Vanlommel & Schildkamp, 2019). Another way for teachers to collect data is data-based decision making (DBDM), in which systematic and deliberate, preset criteria steer the fixed procedures of student-level data collection, analysis, and interpretation (Filderman et al., 2018; Mandinach et al., 2008; Pierce et al., 2013). DBDM includes systematic analysis of existing data sources within schools, collected systematically through structured classroom observation, standardized

19

tests, or formal tests (Hoogland et al., 2016). From the perspective of formative assessment, the focus of DBDM is on using data as a form of feedback to improve teaching and learning (Hoogland et al., 2016). Thus, a huge responsibility lies with the teachers and depends on their knowledge and skills of interpreting the student data and modifying their instruction according to strong data-based decisions, using the available instructional programs and deciding on the intensity of the support for each student (Al Otaiba et al., 2019).

2.2 Assessment and academic skills' development

Assessment is considered a central part of teaching, and it is a key process in effective instruction (Kyttälä et al., 2021; Wiliam, 2011). Therefore, it is a teacher's duty to conduct assessments concerning the students' learning and achievements, and further, to communicate the findings to the students, parents, and the other teachers of the students (Dolin et al., 2018). For daily and long-term instructional decision-making, teachers' judgments about students' academic skills are especially important (Bailey & Drummond, 2006; Begeny et al., 2011; Herppich et al., 2018; Kolovou et al., 2021). A central objective of academic assessment is to support individual learning processes, to guide student learning, and to improve student achievement (Bennett, 2015; Herppich et al., 2018; Kolovou et al., 2021). Teachers monitor the development of the students' academic skills, offer individual support, and make individual instructional decisions along the educational path.

2.2.1 Forms of assessment

Assessment can be seen as two-fold, comprising both the process of demonstrating what students currently know, are able to do, and what kind of assessment supports teaching and learning (formative assessment), and the assessment programs for public reporting and system accountability (summative assessment) (Black & Wiliam, 1998; Looney et al., 2018; Wiliam, 2011). These two forms of assessment differ with respect to assessment purposes (Dolin et al., 2018). When teachers gather data on the students' current skills and learning process and the aim is to foster student learning by providing them more individual instruction, they use formative assessment (Afflerbach, 2016; Dixson & Worrell, 2016; Dolin et al., 2018; Herppich et al., 2018; Marchand & Furrer, 2014). In education, both formative and summative assessments are important in planning for instruction and supporting learning (Stiggins, 2002), and all specific assessment formats can be characterized as either summative or formative (So & Lee, 2011). In the ideal situation, formative and summative assessments complement each other, although they have different purposes. To get a comprehensive view of students' current knowledge and achievement levels, teachers can combine the two forms of assessment, for example, by using their own observations along with standardized tests (Dixson & Worrell, 2016).

In the literature, formative assessment is also called assessment for learning (So & Lee, 2011; Wiliam et al., 2004; Wiliam, 2011), which includes continuous monitoring by the teacher to determine whether the degree of difficulty of the lesson is in correspondence with the students' cognitive and motivational-affective learning processes (Urhahne & Wijnia, 2021; Wiliam, 2011). Formative assessments are frequent, they involve judgments about student performances (quality of responses), and they are used immediately in respect to guiding and improving students' learning (Dixson & Worrell, 2016; Roskos & Neuman, 2012) as well as for adjusting instruction, based on the judgments (Thiede et al., 2018). If teachers need to evaluate students' learning outcomes after completing a certain learning sequence, they use summative assessment (Black & Wiliam, 1998; Cornelius, 2013; Dixson & Worrell, 2016; Dolin et al., 2018; Herppich et al., 2018).

Assessment can be defined as *high-stakes* or *low-stakes* assessment. The former is connected to summative assessment, having important consequences for the student, for instance, how much the student has learned in a certain assessment point, when an instructional segment (e.g., a term) has ended (Dixson & Worrell, 2016; Dolin et al., 2018). Examples of high-stakes assessment are final exams, college entrance exams (e.g., the Scholastic Assessment Test [SAT] in the US), and state-wide or national tests (Dixson & Worrell, 2016). Low-stakes assessment pertains to formative assessment, for example, assessments before and during instruction, in the form of observations, homework, self-evaluations, and Curriculum-Based measurement (CBM) (Dixson & Worrell, 2016).

One form to assess academic skills is *dynamic assessment* (DA), which is an "umbrella term" describing a heterogeneous range of assessment approaches, used in *response to intervention* (RTI) model in the US (Grigorenko, 2009). DA is used for identifying students' learning potential and for predicting their academic skill development (Cho et al., 2020; Grigorenko, 2009). The main characteristic of DA is that the focus is on understanding the student's learning potential and differentiating between "what is it now" and "what can it be in the future" (Cho et al., 2020, p. 720). RTI model consists of three tiers. Tier 1 (primary prevention) includes screening and monitoring of all students, aiming at identifying the at-risk children. In Tier 2 (secondary prevention), the children not responding to given support in Tier 1 are provided with more intensive support, and when students are not responding adequately to the interventions in Tier 2, they are provided with continued and also individual research-based interventions, which is Tier 3 (tertiary prevention). (Björn et al., 2016; Fuchs et al., 2012).

2.2.2 Subject of assessment

Assessment of students' academic skill development in different phases of the school path is essential for teachers to be able to plan instruction and to provide individual support. At school entry, regarding Finnish children's reading acquisition, *phonological awareness* and *letter knowledge* are the main predictors of reading and spelling accuracy (Aro, 2006; Holopainen et al., 2001; Lerkkanen, 2004). In transparent orthographies such as that of Finnish, reading instruction is purely phonics-based instruction focusing on grapheme–phoneme correspondences

21

and phonemic assembly (Aro & Wimmer, 2003). Of the Finnish first grade students, approximately 30% can already decode before starting school, and within the first months of the beginning of reading instruction, the rest of the children also achieve good decoding skills (Lerkkanen et al., 2004). *Decoding* refers to being able to use knowledge of letter–sound correspondence in order to read separate words, and it requires competence in deriving efficient mental representations from the printed text, and also the knowledge of letter patterns, which is essential for correct pronunciation of written words (Ehri, 1994; Grimm et al., 2018).

Phonological awareness is a comprehensive term referring to awareness of large spoken units, such as syllables and rhyming words, and the ability to focus on and manipulate the smallest units of spoken language, phonemes (Ehri et al., 2001). *Phonemic awareness* is connected to a larger construct of phonological awareness and refers specifically to the ability to discriminate and manipulate the sound structure of language, for example, by picking out the initial sound in a word, blending phonemes together, and breaking apart syllables (Anthony & Francis, 2005; Ehri et al., 2001).

Letter knowledge, together with phonemic awareness, has been shown to be the best predictor of learning to read (Ehri et al., 2001; Holopainen et al., 2001; Torppa et al., 2006). Letter knowledge (or alphabet knowledge) is a core emergent literacy skill, referring to a child's knowledge of letter forms, letter names, and corresponding sounds (Piasta et al., 2021). Letter knowledge demands environmental exposure to printed material; however, that does not seem to explain all the variability in adopting letter knowledge (Torppa et al., 2006). Letter knowledge is connected to *spelling and writing* development and the understanding of letter–sound correspondences (Georgiou et al., 2019; Piasta et al., 2021). Acquisition of spelling procedures is affected by the orthographic consistency of the language, features of oral language, and linguistic structure complexity (Holopainen et al., 2017). To spell the word correctly, one has to understand the links between the phonemes and the letters, and be able to use these links in order to spell, especially unfamiliar words (Torppa et al. 2017).

It has been estimated that approximately 16% of people have RD (Catts et al., 2012). Conceptualizing and classifying reading difficulty or specific reading disability itself includes several constructs in the literature. In general, RD can be defined as the outcome of confronting deficits in perception, cognition, or environment (Gaab et al., 2019). One commonly used term for this is dyslexia, which is connected with the medical definition and refers to a specific developmental learning disability or disorder in reading based on neurobiological origin (American Psychiatric Association, 2013; Bigozzi et al., 2015; Landerl et al., 2019; Lyon et al., 2003; Ramus et al., 2018). Dyslexia, in general, refers primarily to a difficulty with accurate and/or fluent word recognition connected with poor spelling and decoding abilities, and reading comprehension difficulty is seen as a consequence of the primary difficulty (Lyon, 2013). Being diagnosed as dyslexic means that an individual having the intelligence, motivation, and schooling to be a good reader has an unexpected difficulty or poor performance in reading (Lyon et al.

2003). Prior studies have confirmed that, especially in the presence of a familial risk for RD, LDs can be rather stable throughout primary school (Dandache et al., 2014; Eklund et al., 2015). For instance, a study by Eklund et al. (2015) found that students with familial risk for RD having *reading fluency* difficulties in Grade 2 still struggled with slow reading speed in Grade 8.

After learning to decode, Finnish students' reading difficulties are usually connected with reading fluency (Aro & Wimmer, 2003; Eklund et al., 2015; Hintikka et al., 2008; Torppa et al., 2015). Dysfluent readers have difficulties in identifying the words in the text since their decoding has not become automatic (Hudson et al., 2008; Kuhn et al., 2010; Meisinger et al., 2010; Wagner & Espin, 2015). Moreover, difficulties in reading fluency can lead to difficulties in understanding the meanings of the words, which can further hamper reading comprehension (Hudson et al., 2008; Mancilla-Martinez & Lesaux, 2010). Reading comprehension includes the complex process of combining various cognitive skills, such as the visual perception of text, ability to hold text in working memory while decoding, construction of mental schemas, and retrieval of prior knowledge from long-term memory storage (Kintsch & Kintsch, 2005; Shapiro et al., 2017). According to the simple view of reading (SVR; Gough & Tunmer, 1986), reading comprehension is seen as the result of, on the one hand, efficient decoding, including quick and accurate reading of isolated words, and on the other hand, linguistic comprehension. The formula of this model represents four reader types: a good reader who can decode and also understand what they read; a poor comprehender who does not understand what they read but who can decode well; a dyslexic reader who has difficulty in word decoding but who can understand what they read; and a poor decoder and poor comprehender who has difficulties in both decoding and comprehension.

Mathematical skills develop hierarchically consisting of basic number knowledge, verbal counting skills, memorizing arithmetic facts, understanding, and using basic arithmetic operations (e.g., addition and subtraction), understanding concepts, and the capability to follow relevant procedures (Butterworth, 2005; Geary et al., 2012; Jordan et al., 2009; Zhang et al., 2020). Students struggling with learning mathematics can be divided into two groups: students having mathematical LDs (MLDs) and students with mild but persistent low achievement in mathematics (Zhang et al., 2020). Arithmetic difficulties manifest themselves in dysfluent retrieval of arithmetic facts and consistent use of counting-based strategies (e.g., solving 1 + 2 by counting "one, two, three") (Koponen et al., 2018). In the elementary grades, math difficulty manifests itself in students' lower performance on counting (Nelson & Powell, 2018), arithmetic fluency, computation, and comparing quantities (Nelson & Powell, 2018; Tolar et al., 2016). Additionally, word problem solving can be difficult for students struggling with math (Kingsdorf & Krawec, 2014).

There is considerable scientific evidence that LDs are heterogeneous and that children often have more than one difficulty (Branum-Martin et al., 2012). LDs have shown strong *comorbidity*, meaning that one academic difficulty often occurs together with some other or even several other difficulties (Cirino et al.,

23

2018; Koponen et al., 2018). This can be seen as a challenge regarding assessment. Common underlying cognitive processes (e.g., working memory and processing speed) involved with the development of reading, spelling, and mathematics skills have been found (Child et al., 2019; Koponen et al., 2018; Watson & Gable, 2013; Willcutt et al., 2019), and the co-occurrence of LDs can present themselves in various ways. For instance, a student with RD can have additional difficulty in spelling (Holopainen, Kiuru et al., 2017; Landerl & Moll, 2010; Landerl & Wimmer, 2008), language development (Adlof & Hogan, 2018; Spanoudis et al., 2019), or mathematics (Child et al., 2019; Koponen et al., 2018; Moll et al., 2019), more frequently than can be expected at random.

Overlapping LDs in reading and math are usually more severe as compared with difficulties occurring in isolation (Branum-Martin et al., 2012; Fletcher et al., 2007). In the cases where LDs overlap, exact assessment of the core difficulty may be rather challenging, and despite provided support, the skill development can turn out to be slow, and the students can lag constantly behind the typical age level (Vaughn & Wanzek, 2014; Wanzek et al., 2014). This can lead to impairing the students' academic careers until the end of compulsory education and beyond (Eklund et al., 2015; Eklund et al., 2018; Holopainen, Taipale et al., 2017; Koponen et al., 2018).

Teachers' monitoring of the development of the students' academic skills across primary school is important, not only for the students themselves, but also regarding the special education support resources. They are generally limited, and after identifying the LD, the effect of the given support should be demonstrated and modified or ended, when necessary. For the positive development of the students' academic skills, prevention of LDs, early recognition of the need for support, and provision of support as soon as possible are crucial. Teachers' judgments of the students' academic skills are of significant importance when making decisions about students' instruction and individual support, aiming at improving student learning (Herppich et al., 2018; Kolovou et al., 2021). However, academic skills develop individually, and being able to provide relevant support calls for thorough, continuous, and systematic assessment procedures (Peterson et al., 2017).

2.3 Teachers' assessment of literacy and math

Teachers' assessment practices can be divided into three main categories: *tests* (individual or screening test batteries), *qualitative assessments*, and *CBM assessments*. Qualitative, student performance-based assessments focus on the observation of the students' learning and development, aiming to complement the information from tests or exams (Campbell, 2001). They generally include teachers' own observations, rating scales, and discussions (Atjonen, 2014; Bailey & Drummond, 2006; Shapiro et al., 2017; Südkamp et al., 2012). CBM is a progress monitoring system used in tracking the progress of an individual student with learn-

ing difficulties, enabling the evaluation of the effectiveness of provided instruction (Deno, 1985, 2003). CBM makes it possible for the teachers to determine the students' levels of performance, and to see their rate of progress toward certain learning goals in a valid and reliable manner (Deno, 2003). For example, a formative curriculum-based assessment of reading (CBM-R) within the RTI framework is commonly used as a universal screening tool, given several times during each grade, to identify the students at risk for failing to reach the reading standards set for the grade level (Marchand & Furrer, 2014).

Teachers may also assess students' academic performance using either direct (e.g., achievement tests) or indirect measures (e.g., their own observations) (Begeny et al., 2011; Feinberg & Shapiro, 2003; Woolley, 2008). To assist in the identification of RD, teachers can use rating scales that can focus on specific literacy skills (e.g., letter knowledge) or general reading ability (Titley et al., 2014). Nevertheless, a key weakness of rating scales is that there are usually no clear cut-off scores for discrimination of students performing typically and the at-risk students (Titley et al., 2014). In Finland, currently only one nationally representative standardized test for assessing difficulties in reading (phonological awareness, technical literacy, reading fluency, and reading comprehension) at primary school is available, the Ala-asteen lukutesti (ALLU; Reading Test Battery for Primary School) (Lindeman, 1998). In addition, SEN teachers often use self-designed tools and materials to identify the risk for RD (Soodla et al., 2021; Takala & Ahl, 2014). To be able to gain an overall understanding of the actual level of the student's reading skills and to make relevant instructional decisions, teachers need to assess reading accuracy, reading fluency, and reading comprehension. And in cases where a student is manifesting comorbid and persistent LDs, it is important to pay attention to assessing and identifying all difficulties.

At the beginning of the child's school path, teachers have the best opportunities to identify the first signs that a child is struggling with reading and what kind of support is needed (Bailey & Drummond, 2006; Begeny et al., 2011). It has been indicated that a child having slow development in letter knowledge and phonological awareness at preschool age (Catts et al., 2009) or at the beginning of the first grade (Lyytinen et al., 2006) can be at risk for RD. Letter knowledge has been found in several prior studies to predict later literacy skills, especially in transparent orthographies (Holopainen et al., 2001; Leppänen et al., 2008; Lerkkanen et al., 2004). In Finland, at the beginning of school, teachers' reading assessment includes evaluation of the students' letter knowledge with letter naming and letter writing tasks, and phonological awareness, generally assessed with phoneme identification, phoneme blending, and phoneme deletion tasks (e.g., Lerkkanen et al., 2006; Poskiparta et al., 1994).

The generic definition of adequate *reading fluency* includes accuracy, rate, and prosody, and in the assessment of fluency, usually the focus is on reading speed and accuracy, operationalized as the number of correctly read items within a certain time limit (Hudson et al., 2008; Kuhn et al., 2010). *Oral* reading fluency can be assessed using read-aloud texts with short passages (Marchand & Furrer, 2014), and *silent* reading fluency can be assessed with maze tasks, where students

25

are asked to read a text and they have to choose the correct word to complete the sentence from given options (Denton et al., 2011; Kim et al., 2015), and word chain (strings of words without separating spaces) and sentence verification tests (Denton et al., 2011). It has been indicated that when evaluating students' oral reading fluency, teachers have faced difficulties in rating the students' reading levels in certain categories (low, average, or high) (Begeny et al., 2008). For example, the frequently used ALLU test (Lindeman, 1998) in Finnish reading research includes time-limited word recognition tasks, and a student has to connect the words with the correct picture.

Compared with reading fluency assessment, reading comprehension assessment can be rather challenging or even controversial for the teachers as it is connected to the students' language skills and vocabulary, previous knowledge of the topic, reading strategies, and response format (Collins et al., 2017; Tong & Deacon, 2017). Reading comprehension can be assessed by silent reading tasks with text-related multiple-choice questions (Oslund et al., 2018), picture matching tasks (García & Cain, 2014), maze tasks (Marchand & Furrer, 2014), or oral or written retell tasks (Bernfeld et al., 2013; Reed & Vaughn, 2012). For example, with the Finnish ALLU test (Lindeman, 1998), reading comprehension can be assessed using factual or story texts, with multiple-choice questions connected with them. Currently, this test is also available in a digital form (Centre for Learning Research, University of Turku).

Concerning *math* assessment, there are certain mathematical skills that students are expected to gain, according to the curriculum, in a certain grade or by a certain age that also guide the assessment. For example, in Grades 1 and 2, a Finnish child should be able to name and write large whole numbers, count forwards, backwards, or skip count, identify place value of digits, do simple addition and subtraction tasks, and retrieve some arithmetical facts from memory (FNBE, 2014; Mononen & Aunio, 2013). Later, students in Grades 3 to 9, for example, are expected to be able to master basic algebraic concepts, percentages, and ratios, as well as understand fractions (FNBE, 2014; Soares et al., 2018). In Finland, basic arithmetic skills can be assessed, for instance, with individual game-like assessment tasks or paper-pencil group tests that have a time limit (Koponen et al., 2018).

According to prior research, in addition to appearing in the early school years, LDs might emerge only later, after the basic reading and spelling skills have been adopted, and they can turn out to be long-lasting (Etmanskie et al., 2016; Kent et al., 2019; Oslund et al., 2018; Torppa et al., 2017; Torppa et al., 2015; Wanzek & Vaughn, 2009). However, identification of LDs at that phase can be more challenging for the teachers, caused by the ineffectiveness of the screening tools that have been designed for younger students (Etmanskie et al., 2016; Leach et al., 2003). In addition, providing support for manifested learning challenges can become more complicated because of the students' overlapping learning and behavioral or attention difficulties. Also, it has been indicated that inattentive behavior in particular may stabilize only in the later phase of childhood (Lahey et al., 2005) and that inattentive behavior in Grade 1 can predict reading difficulties

in Grade 3 (Morgan et al., 2008) or later in adolescence (Massetti et al., 2008). Therefore, after the early grades of primary school, follow-up on the development of students' LDs is vital.

As noted before, overlapping difficulties in learning can slow down students' academic development, and when the difficulties are severe, it is of utmost importance that the difficulties are carefully investigated and individual support is implemented accordingly. To avoid inefficient instructional practices and to demonstrate the students' response to an intervention, the progress of the academic skills development should be regularly and frequently monitored (Björn et al., 2018). For example, in the RTI framework, the students are monitored systematically every week during the five to eight weeks of intervention, even in the primary prevention phase, while in Finland, the progress of students receiving general support is monitored a maximum twice per year (Björn et al., 2018).

2.4 Accuracy of teachers' assessment

As teacher judgments are the primary source of information about student achievement, being able to make accurate assessments is considered a key aspect of teachers' professional competence (Meissel et al., 2017). Accuracy of teachers' assessments or judgments can be defined as the teachers' ability to make adequate assessments of student characteristics and appropriate estimations of learning and task requirements (Artelt & Gräsel, 2009). Further, judgment accuracy can be operationalized as a correlation between the teacher judgments and students' test performance (Kolovou et al., 2021) or to the extent to which teachers' judgments correlate with some other evaluation criteria (Kaufmann, 2020; Südkamp et al., 2012; Südkamp et al., 2018; Thiede et al., 2018). Absolute judgment accuracy refers to the extent to which the judgment demonstrates overconfidence (judgment is more optimistic than actual performance) versus underconfidence (judgment is less optimistic than actual performance) (Dunlosky & Rawson, 2012; Thiede et al., 2018). Relative judgment accuracy stands for "the degree to which judgments discriminate between different levels of performance across students," according to Thiede et al. (2018, p. 107), or it is described as "the intra-teacher correlation between predicted and actual performance computed across students" (p. 107). It has been shown that teachers tend to overestimate their students' achievement, which can result in not identifying the at-risk students or students in need of support. From the perspective of judgment accuracy, this can be considered mostly a disadvantage, meaning that teachers may disregard students' lack of effort, test anxiety, and forgetting of subject matter (Urhahne & Wijnia, 2021).

According to a recent review on the accuracy of teacher judgments (Urhahne & Wijnia, 2021), there are three kinds of teacher judgment accuracy types: person-related, task-related, and person-specific (Karst et al., 2014). Person-related judgment accuracy is connected with the teacher's knowledge of the student characteristics (academic achievement, cognitive abilities, academic self-

27

concept, learning motivation, work, and social behaviors) affecting the formative and summative assessment of educational progress, and most of the current research on the accuracy of teacher judgments covers this person-related component (Urhahne & Wijnia, 2021). A situation where the teacher has to assess the difficulty of the tasks based on the students' performance level requires task-related judgment accuracy, and person-specific judgment accuracy is related to the teacher's capability of identifying and classifying individual students' personal characteristics or behaviors (Urhahne & Wijnia, 2021). The difference between person-related and person-specific judgments is that the latter judgments are aimed at identifying exceptional students, instead of assessing each student in relation to a certain characteristic (Urhahne & Wijnia, 2021).

By adapting a *sociocultural view to assessment literacy*, assessment literacy can be conceptualized as a social practice dependent on dynamic context (Willis et al., 2013). It has also been found that personal thoughts and perceptions can have a mediational role in teachers' evaluations of the students' skills, and their judgments may also be related to some personal characteristics, such as their own skills, training, future expectations, or perceptions (Kikas et al., 2015; Soodla & Kikas, 2010). When we take all these matters into account, it becomes obvious that teachers' decision-making is often a process where teachers have to balance the demands caused by external factors with their own values and beliefs (McMillan, 2003; Xu & Brown, 2016).

Earlier studies have shown that teachers' ratings of average- or high- performing student's reading skills are generally in accordance with standardized achievement test scores (Bailey & Drummond, 2006; Begeny et al., 2011; Shapiro et al., 2017; Südkamp et al., 2012). The median correlation in a study by Hoge and Coladarci (1989) between teacher judgments and students' test scores was .66. In a recent reanalysis of the same data of the aforementioned meta-analysis indicated that in the antecedent study, teachers' judgment accuracy had been underestimated, and the correlation reported in the new reanalysis was .80 (Kaufmann, 2020). Kaufmann (2020) also suggested that the same kind of underestimation had occurred in another meta-analysis by Südkamp et al. (2012), where a median correlation of .53 was found. Kaufman's (2020) results demonstrated that teachers' judgment accuracy is actually higher than the previous studies had indicated. However, successful early identification of at-risk and not at-risk students requires universal screening procedures in addition to teacher judgment (Compton et al., 2010; Martin & Shapiro, 2011; Snowling et al., 2011). The satisfactory accuracy level of a screening measure is connected to how correctly the students with RD and typically achieving students are identified (Ball & O'Connor, 2016). Of these two aspects, sensitivity (accuracy of identifying the struggling students) has usually been considered more important so that the students most in need are also provided support. In contrast, specificity refers to how accurately the students not at risk are identified (Ball & O'Connor, 2016). From the perspective of targeting learning support resources to the students that need support the most, the high level of specificity can be considered preferable. An acceptable level of

classification accuracy for sensitivity is considered 90% or above, and for specificity, it should be 80% at least (Catts et al., 2015; Compton et al., 2010).

It is well known that when the teachers' assessments are task-specific, as opposed to general assessment, it enhances the reliability of the assessments (Feinberg & Shapiro, 2009; Titley et al., 2014). To gain better accuracy in the teachers' assessments, using specific rating scales for reading fluency and reading comprehension that are in line with the test batteries could improve the accuracy of the assessments (Kikas et al., 2017; Paleczek et al., 2015, 2017). Nevertheless, teachers' assessment of reading fluency and reading comprehension has proved quite inaccurate: correlations between the teacher ratings and the test scores have mostly been moderate (.50 to .60) in Grades 1 and 3 (Martin & Shapiro, 2011; Paleczek et al., 2015, 2017). However, in contrast to most of the prior results, Fedora (2015) found in her study that teachers were more accurate in identifying the first grade students struggling in reading as compared with those achieving typically.

In comparison to teacher ratings, which tend to have high false-negative and low true positive rates, standardized tests can more accurately identify students at risk for RD (Fletcher & Satz, 1984; Flynn & Rahbar, 1998). Therefore, using a combination of various screening measures can lead to the most accurate outcome (Compton et al., 2010; Flynn & Rahbar, 1998; Martin & Shapiro, 2011; Snowling et al., 2011; Speece et al., 2011). The assessments should also be repeated frequently, but similar kinds of tests should be alternated, which decreases the risk that a student can improve the scores simply due to practice (Duff & Clarke, 2011). The accuracy of teachers' assessments can also depend on the nature of the assessment tool (universal or task-specific) and the situation of the assessment, and teachers are recommended to use the assessment practices that are best fitted for pedagogical decisions (Karst et al., 2018).

2.5 The aims of the thesis

The main aim of this thesis was two-fold: firstly, to investigate primary teachers' assessment practices of students' academic skills, and the accuracy of the assessments. And secondly, to study the relation between the reasons and the length of part-time special education to the development of the students' academic skills (reading fluency, reading comprehension, and arithmetic fluency).

The more specific research aims of the present thesis (see Figure 1) were the following:

1. To investigate teachers' assessment practices of pre-reading skills to identify the risk of RD in Grade 1 (class teachers and SEN teachers), and difficulties in reading fluency and reading comprehension in Grade 6 (SEN teachers). Additionally, to examine the accuracy (sensitivity and specificity) of the teacher ratings, by investigating the associations between the teacher ratings and the students' test scores (Sub-studies I and II).

2. To study the SEN teachers' reported reasons and additional difficulties of the Grade 1 students receiving part-time special education; how the students' skills (reading fluency, reading comprehension, and arithmetic fluency) in two support groups differed as compared with each other and the control group in Grade 1 and Grade 4; and how the lengths of support periods were related to the development of reading and arithmetic skills across Grades 1 to 4 in the two support groups as compared with the control group (Sub-study III).

The first sub-study addressed the use of both class teachers' and SEN teachers' assessment practices and how accurately the students' risk for RD or typical development of pre-reading skills had been identified in Grade 1 based on the teachers' ratings and the students' test scores. Sub-study II concentrated on how important the SEN teachers ranked reading assessment practices, and how accurately they had been able to identify the Grade 6 students' difficulties in reading fluency or reading comprehension, or typical performance in those skills. In the third sub-study, two groups of students receiving part-time special education were followed up from Grade 1 to Grade 4. The students were divided into two support groups based on the length of part-time special education, and their peers in the large follow-up sample formed the control group. The differences in reading fluency, reading comprehension, and arithmetic fluency in Grades 1 and 4 between the support groups and the control group were investigated. Also, the development of the reading and arithmetic skills across Grades 1 to 4 in the two support groups was reflected against the reasons for support, additional learning-related difficulties, and the development of the control group.

Research question 1: Sub-study I What 1) Which assessment practices do class teachers and SEN teachers use to assess pre-reading assessment skills? Specific research practices do 2) Are teacher ratings associated with test scores of pre-reading skills? aims: teachers use in 3) How accurately do teachers identify students' pre-reading difficulties as compared with test identification of scores, and what are the sensitivity and specificity rates of their assessments? RD and The main aim of this difficulties in thesis was 1) to reading, and investigate teachers' Sub-study II how accurate assessment practices of are the 1) What kinds of assessment practices do SEN teachers use to evaluate students' reading students' academic assessments performance, and how do they rank different practices? skills in primary school, compared with 2) To what extent are SEN teachers' ratings of Grade 6 students' reading fluency and reading the correlations between students' test comprehension skills associated with students' test scores for the same skills? the teacher ratings and scores? 3) How accurate are the SEN teachers' perceptions of their students' reading fluency and the test scores, the reading comprehension skill levels, as compared with the students' test scores? accuracy of the Research assessments, and 2) to study the reasons for question 2: Sub-study III part-time special What are the education and the effect reasons for of part-time special 1a) What reasons did SEN teachers report for providing part-time special education in Grade 1, part-time education on the and to what extent did the reasons for support differ between the Early-stage support (ESS) and special development of Long-lasting support (LLS) groups? education, what students' academic 1b) What kinds of additional learning-related difficulties did SEN teachers report in Grade 1, kind of skills. and to what extent did the ESS and LLS groups differ in terms of the extent and characteristics additional of those difficulties? difficulties do 2) Did the two support groups differ in reading fluency, reading comprehension, and arithmetic students have, fluency at the beginning of part-time special education in Grade 1 as compared with each other and how is the and their peers in the large follow-up sample (the control group)? length of 3) Did development of reading fluency, reading comprehension, and arithmetic fluency in the support period ESS and LLS groups across Grades 1 to 4 reflect the duration of part-time special education, and related to to what extent did the ESS group, the LLS group, and the control group differ in terms of those development of skills at the end of the follow-up? academic skills?

Research questions in the original sub-studies:

FIGURE 1 Specific research aims and research questions of the thesis.

3 METHOD

3.1 Participants and data

The samples of all three sub-studies were drawn from two Finnish longitudinal studies: Jyväskylä Longitudinal Study of Dyslexia ([JLD]; Lyytinen et al., 2006) and the First Steps follow-up study (Lerkkanen et al., 2006–2016). All the participants in the teacher samples were working in mainstream schools, and participating students were attending general education. The samples from both longitudinal studies were representative of the Finnish population with respect to maternal education distribution (OSF, 2019).

In sub-study I, the focus was on the data collected in the fall term of Grade 1 (the JLD data were collected during 2000–2003, and the First Steps study data in 2007); in sub-study II, in the spring term of Grade 6 (First Steps follow-up study, collected in the spring term of 2013); and in sub-study III, in the fall and spring terms of Grade 1 and in the spring terms of Grades 2, 3, and 4 (First Steps follow-up study, collected 2007–2008, 2009, 2010, and 2011, respectively).

The number of teachers varied in the three sub-studies, and class teachers and SEN teachers participating in the JLD study were involved only in sub-study I. No demographic background information was available for the teachers in the JLD study. In sub-study I majority of the class teachers in the First Steps study (78%) had a master's degree in education from a class teacher program (69%), from a special education program (5%), or from both programs (4%). One percent of the class teachers had a doctoral degree in education. Twenty-one percent had some other degree, typically a bachelor's (BA) degree, which was formerly sufficient for the qualification as a class teacher or kindergarten teacher. In sub-studies I and III the SEN teachers had a master's degree from a teacher education program combined with a SEN teacher qualification (53%), a master's degree from a SEN teacher program (44%), or some other certification (3%), such as a SEN teacher qualification combined with a BA degree as a kindergarten teacher. Two class teachers and two SEN teachers did not report their education. In sub-study

II 97% of the SEN teachers had a masters' degree, from a class teacher program combined with a SEN teacher qualification (52%) or from a SEN teacher program (45%). Three percent had another basic education such as a kindergarten teacher combined with a SEN teacher qualification. Only teachers who had completely and carefully filled in the individual observation forms of the students participating in the follow-up studies were selected for the sub-studies. During the two follow-up studies, the students participated yearly in the individual or group tests administered by trained testers. The teacher and student samples in the three sub-studies are presented in Figures 2, 3, and 4.

Jyväskylä Longitudinal Study of Dyslexia. The JLD study focused on investigating familial risk dyslexia, and these data were used in sub-study I. Two hundred children from Central Finland were followed up from the children's birth until they were teenagers. The children's data (N = 200) comprised four successive age cohorts of first grade children (M age = 7.19 years, SD = 0.26; 53% boys) born from 1993-1996, of which half had a familial risk of dyslexia, and the other half represented the control group. The at-risk children were defined by parents' self-reports and test data indicating literacy difficulties and their reports of similar problems among their immediate relatives. The questionnaire sent to the parents dealt with demographic information and the incidence of reading and spelling difficulties in the parents' childhood and adulthood and among their relatives. In the diagnostic tasks of reading and spelling, the parents were selected in the dyslexic sample if they scored below -1 z-scores in either spelling accuracy or speed, or accuracy of oral text reading, and also in two or more out of eight computer-aided measures. The children selected for the control group or low-risk group did not have any reported parental literacy difficulties or literacy difficulties in their first- or second-degree relatives (for more specific details, see Leinonen et al., 2001). For this thesis, all teachers who had filled in and returned the individual student observation forms and the ratings of the students' prereading skills (letter knowledge and phonological skills) were included. The number of students included in sub-study I was determined by the fact that the studied skills had been rated by the teachers and the test scores of the same skills were available (class teachers' ratings of letter knowledge, n = 44; phoneme identification, n = 42; and phoneme blending, n = 42 students; and SEN teachers' ratings of letter knowledge, n = 36; phoneme identification, n = 33; and phoneme blending, n = 35 students).

The First Steps study. In all three sub-studies, data from a comprehensive First Steps follow-up study, including data on approximately 2,000 children and their parents and teachers, were used. The participants of the First Steps study consisted of one whole age cohort of children in three participating municipalities and half of the age cohort in one municipality. Parental education levels in the data set followed the Finnish national average (European Commission, 2013). The children's teachers and their parents gave informed consent for participation in the study. From the total follow-up sample of children, the children's risk for RD was determined by the researchers at the end of the kindergarten year on the

basis of four criteria: children's initial phoneme identification (indicator of phonological awareness), letter knowledge, rapid automatized naming, and parental report of their own RD in childhood or adulthood (see Lerkkanen et al., 2011). The risk for RD was defined as a joint occurrence of at least two criteria out of three: low phonological awareness (i.e., scored clearly below the age level in initial phoneme identification, ≤15th percentile); poor letter knowledge (≤15th percentile); and poor rapid automatized naming (≤15th percentile) (Kiuru et al., 2013; Lerkkanen et al., 2011). In cases where RD was reported by the parents, the factor for being identified as being at risk for RD, a score at or below the 15th percentile in one of the three tests (phonological awareness, letter knowledge, or rapid automatized naming) was adequate. The control group children were randomly selected from the same classrooms as the children identified as being at risk for RD. A subsample of 598 children from the total study sample (of which half had a risk of RD and the other half represented the control group) was selected for more intensive follow-up by the researchers. Of this subsample, only those students who had been rated using the individual student observation form by the SEN teachers and whose test scores were available were selected for the sub-study samples (Sub-study I, n = 69, and sub-study II, n = 55). In sub-study III, 130 students having received part-time special education and whose test scores were available were selected for the two support groups, and the peers (n = 2,027) in the follow-up study belonged to the control group.

Sub-study I (class teachers and SEN teachers): The data from this sub-study were drawn from the JLD study and the First Steps study (Figure 2). The teacher samples comprised class teachers from the JLD study (n = 16) and the First Steps study (n = 24), and SEN teachers participating in the JLD study (n = 29) and the First Steps study (n = 34). The assessment practices of the class teachers participating in the First Steps study were reported with the SEN teachers' student observation form, and the class teacher data included only their assessment practices, without the student ratings. For the ratings, the teachers filled in the individual student observation forms. The student samples from the JLD study (n = 33-44) and the First Steps study (n = 69), respectively, comprised the students rated by the class teachers and the SEN teachers (JLD study), or by the SEN teachers only (First Steps study), and whose reading test scores from Grade 1 fall were available (See Figure 2).

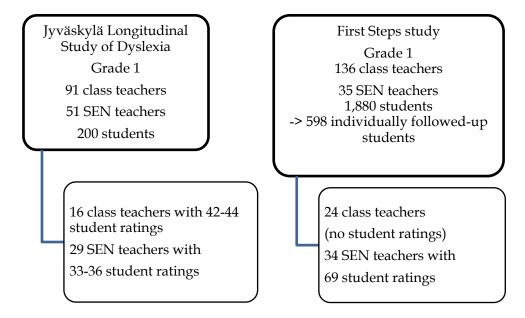


FIGURE 2 Samples in sub-study I.

Sub-study II (SEN teachers): In this sub-study, the First Steps study data collected during Grade 6 in the spring of 2013 were used. The SEN teachers were sent a questionnaire and a list of students who were individually followed up in their school, and they were asked to rate 1–6 students who had received part-time special education in Grade 6, using an individual student observation form. Twenty-nine SEN teachers returned the questionnaires, and they rated 55 students altogether. The students were among the aforementioned more intensively followed subsample whose level of reading fluency and reading comprehension had been rated by the SEN teachers and whose test scores of reading fluency and reading comprehension were available. All selected students' risk for RD had already been defined at the end of the kindergarten year, they had received part-time special education in either reading and/or mathematics during the previous data collection point in Grade 4, and they all had received part-time special education in Grade 6 (see Figure 3).

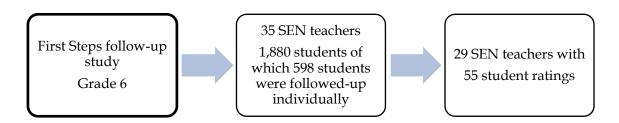


FIGURE 3 Samples in sub-study II.

Sub-study III (SEN teachers): The samples of this sub-study were drawn from the First Steps follow-up study. The SEN teachers (n = 35) had evaluated the students' need for part-time special education in Grade 1 by responding to an open-

ended question: "What is the main reason for support?" The students in the sample selected for this sub-study (n = 130) had received special education during Grade 1. They were participating in the First Steps study and they belonged to the more intensively followed-up subsample of the follow-up study. One criterion, in addition to having attended part-time special education in Grade 1, to be selected in the student sample was that the student had to have the test scores from Grades 1 to 4 of reading fluency, reading comprehension, and arithmetic fluency tasks. The 130 students were further divided into two support groups based on the length of part-time special education. The students that had received part-time special education in Grade 1 or in Grades 1 and 2 formed the Early-stage support (ESS, n = 55) group, and the Long-lasting support (LLS, n = 75) group comprised the students whose support had lasted from Grade 1 until Grade 3 or 4. The peers participating in the follow-up study but not involved in the two support samples of this sub-study formed the control group (n = 2,027) (See Figure 4).

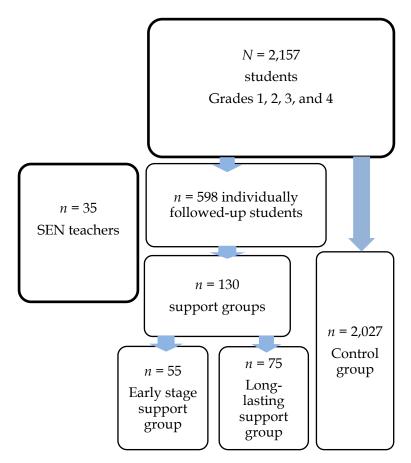


FIGURE 4 Samples in sub-study III.

3.2 Measures

JLD study. The SEN teachers filled in individual student observation forms where they also reported their assessment practices of teaching reading and writing in Grade 1 by responding to an open-ended question: "What kinds of practices have you used in the assessment of learning to read and write, and what sub-skills have you assessed?" Class teachers were not asked about their assessment practices. Both the class teachers and the SEN teachers filled in the observation forms using a 5-point rating scale (ranging from 1 = clear problem to 5 = masters the skill very well) for evaluation of the students' pre-reading skills. For the analyses, the five options were pooled into two categories: Clear problem or mild problem and No problem. Students' letter knowledge was tested with an individual letter naming task in Grade 1, in August, including 29 uppercase letters shown on a sheet of paper in a fixed order, which was shown to the student by a trained tester. The instruction to the student was to name the letters as quickly and accurately as possible. The score was the number of correct responses. Phonological awareness was tested with two individual tests administered by trained testers: the phoneme identification task and the phoneme blending task from the Diagnostic Test Battery (Poskiparta et al., 1994). In the phoneme identification task, the tester first said the words (N = 10) to the student, and the student was then instructed to say which sound was at the beginning of each word. In the phoneme blending task, the tester first said 10 word items, phoneme by phoneme, and the instruction to the student was to say what words were formed. The sum scores were based on the number of correct items. Cronbach's alpha for the phoneme identification task was 0.94 and 0.57 for the phoneme blending task. The variable of phonological awareness was the mean of the standardized sum scores of the two tasks, of which Cronbach's alpha was 0.80.

First Steps study. The class teachers and the SEN teachers gave their background information (age, work experience, job description, etc.) in the teacher questionnaires. Sub-study I: in Grade 1, the teachers reported their assessment practices in the individual student observation forms sent to the SEN teachers by responding to the question: "How was the need for special education defined? What/how/when assessed?" The forms included both class teachers and SEN teachers' responses. Also, the SEN teacher ratings of the Grade 1 students' prereading skills (letter knowledge and phonological skills) were included in substudy I. The ratings were part of the individual student observation form, and the teachers were asked to rate the students' skills on a 3-point scale (1 = clear problem, 2 = mild problem, and 3 = no problem) that was further dichotomized for the analyses to two options (Clear problem/mild problem or No problem). In sub-study II, the SEN teachers were asked to arrange six given assessment practices of reading and writing by marking 1 (the most important), 2 (the second most important), and so on, based on their experienced relevance. The SEN teachers also rated the Grade 6 students' reading fluency and reading comprehension

skills using the same 3-point scale, and for the analyses, those responses were dichotomized into two options (Clear/mild problem or No problem).

The students' pre-reading skills (Sub-study I, fall term of Grade 1), reading fluency, and reading comprehension (Sub-study II, spring term of Grade 6, and sub-study III, spring term of Grades 1, 2, 3, and 4), and arithmetic fluency (Substudy III, spring term of Grades 1, 2, 3, and 4) were tested by trained research assistants in the classrooms with group tests (see Table 1). In Grade 1 (August), letter knowledge was tested with the letter knowledge task (ARMI test battery; Lerkkanen et al., 2006) as an individual test. It consisted of 29 uppercase letters on a sheet of paper that were shown to the student in a fixed order, and the student was given instructions to name them. The score was the number of correct responses (maximum 29). Cronbach's alpha for the letter knowledge task was 0.92. Phonological skills were tested in Grade 1 (August) using a group-administered phoneme blending task (Poskiparta et al., 1994). The experimenter first said nine words, phoneme by phoneme, and after that, the students were shown four pictures of objects from which they had to choose the matching picture for the word they had heard. The score was the sum of correct items (maximum 9). Cronbach's alpha was 0.70.

Reading fluency was tested in Grades 1, 2, 3, and 4 with two group-administered tests. A subtest of the standardized Finnish reading test battery for primary school (ALLU, Lindeman, 1998) is a time-limited word recognition task consisting of 80 pictures with four phonologically similar words. Students were instructed to read the words silently and then connect each picture with the correct word by drawing a line. The time limit was two minutes, and the score was the number of correct answers (maximum 80). The other reading fluency task was a Finnish adaptation of the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner et al., 2009; Finnish adaptation by Lerkkanen & Poikkeus, 2009). The task included 60 sentences, and the students were instructed to first read the sentences silently, decide whether the content of the sentence was true or false, and then circle the appropriate choice. Cronbach's alphas for Grades 1, 2, 3, and 4 were 0.86, 0.82, 0.82, and 0.83, respectively.

In Grade 6, reading fluency was tested using two group-administered tasks. The time-limited word recognition task is included in the Finnish reading test battery for primary school (ALLU, Lindeman, 1998), and the sentence reading fluency task, *Luksu*, is a Finnish adaptation of the Salzburger Lese-Screening (SLS) test (Mayringer & Wimmer, 2003). The first time-limited word recognition task consisted of 80 pictures with four phonologically similar words. Students were instructed to read the words silently and then connect each picture with the correct word by drawing a line between the two. The time limit was two minutes, and the score was the number of correct answers (maximum 80). In the sentence reading fluency task, the students were instructed to silently read a sentence as quickly as possible and then decide whether the statement was true or false. The time limit was two minutes, and the mean of the standardized scores of the two tasks represented the fluency score. Cronbach's alpha for the two tasks was .64.

Reading comprehension was tested with a group-administered subtest of the Finnish reading test battery for primary school (ALLU, Lindeman, 1998) in Grades 1, 2, 3, 4, and 6. It consisted of a short text and 12 multiple-choice questions, each having four answer options from which the students could choose. Students were instructed to silently read the text before answering the questions, consisting of either multiple-choice questions about the text with four possible responses or questions requiring the student to arrange the given statements by their order of appearance in the text. The students completed the task at their own pace without any time limit, and the score was the number of correct answers. Cronbach's alphas for Grades 1, 2, 3, and 4 were 0.98, 0.94, 0.96, and 0.85, respectively, and 0.66 in Grade 6.

Arithmetic fluency was tested in Grades 1, 2, and 3 using the Basic Arithmetic Test and the Basic Arithmetic Test II in Grade 4 (Aunola & Räsänen, 2007) as a group-administered test. Both versions have 28 items. The items in the Grades 1, 2, and 3 version include addition and subtraction, and the items in the Grade 4 version includes addition, subtraction, division, and multiplication. The students were told to do as many of the counting assignments as they were able within a three-minute time limit. The scores were the number of correct responses (maximum 28). Cronbach's alphas for Grades 1, 2, 3, and 4 were 0.98, 0.86, 0.68, and 0.85, respectively. Test scores were further standardized within grade levels to enable the evaluation of progress in the two support groups as compared with the control group.

TABLE 1 Measures in sub-studies I-III.

Sub-			
study			
I	JLD	First Steps study	
	SEN teachers' questionnaires	SEN teachers' questionnaires includ-	
	Class teachers' student ratings	ing class teachers' reported assess-	
	SEN teachers' student ratings	ment practices	
	Letter naming task (individual	Class teachers' questionnaires	
	test designed for JLD)	SEN teachers' student ratings	
	Phoneme identification task (individ-	Letter naming task (individual)	
	ual)	- ARMI test battery (Lerkkanen et	
	- Diagnostic Tests 1, Reading and	al., 2006)	
	Writing (Poskiparta et al., 1994)	Phoneme blending task (group)	
	Phoneme blending task	- Diagnostic Tests 1, Reading and	
	- Diagnostic Tests 1, Reading and	Writing (Poskiparta et al., 1994)	
	Writing (Poskiparta et al., 1994)		
II	First Steps study		
	Reading fluency tasks (group)		
	- Finnish reading test battery for primary school (ALLU; Lindeman, 1998)		
	- Luksu (Finnish adaptation of the SLS test; Mayringer & Wimmer, 2003)		
	Reading comprehension task (group)		
	- Finnish reading test battery for primary school (ALLU, Lindeman, 1998)		

III	First Steps study		
	SEN teachers' student ratings, Grade 1 Reading fluency tasks (Grades 1, 2, 3, and 4) (group)		
	- Finnish reading test battery for primary school (ALLU TL2; Lindeman,		
	1998)		
	- TOSREC (Wagner et al., 2009)/Finnish adaptation (Lerkkanen & Poikkeus,		
	2009)		
	Reading comprehension tasks (Grades 1, 2, 3, and 4) (group)		
	- Finnish reading test battery for primary school (ALLU; Lindeman, 1998)		
	Arithmetic fluency tasks (Grades 1, 2, 3, and 4) (group)		
	- Basic Arithmetic Test, and Basic Arithmetic Test II (Aunola & Räsänen,		
	2007)		

3.3 Analytical strategies

In all sub-studies, IBM's SPSS program (versions 24 and 26) was used. In *sub-studies I and II*, descriptive statistics (frequencies, percentages, means, and standard deviations) were used to demonstrate the teacher-reported assessment practices. To show the associations between the teacher ratings and the test scores, the non-parametric Spearman's rank order correlation coefficient was used because the data were not normally distributed. In sub-study I, a two-tailed test and in sub-study II, a one-tailed test was used. To analyze the accuracy of the teacher ratings, binary logistic regression analyses were conducted. Before the statistical testing, the variables were dichotomized and the test scores were used as dependent variables and the teacher ratings as independent variables. In sub-study I, the cut-off score of students' low performance in the test was set to the 15th percentile, and in sub-study II, to the 16th percentile, based on the large First Steps study sample in Grades 1 and 6.

In sub-study III, descriptive statistics were used to specify students' teacher-reported LDs and additional learning-related difficulties, with cross-tabulations and χ^2 tests for identifying differences between the two support groups. The magnitude of possible differences in the three investigated skills among the groups (ESS, LLS, and control groups) was investigated with Cohen's d (Cohen, 1992) by calculating the effect sizes (0.20 small, 0.50 medium, and 0.80 large). The differences in skill development among the three groups were examined using mixed-design Analyses of Variance (ANOVAs). Grade levels 1, 2, 3, and 4 were used as the within-subjects factor, and the group (ESS, LLS, and control) was the between-subjects factor. Standardized scores of the academic skills enabled showing relative differences between the support groups as compared with the control group, and also group comparisons of the reading fluency, reading comprehension, and arithmetic fluency skills. The main effect of the grade was not reported, for the reason that standardization was performed separately for each grade and measure.

TABLE 2 Overview of the themes, variables, and statistical methods in sub-studies I–III.

Themes in sub-studies	Variables	Statistical methods
Sub-study I: Teachers' assessment practices of pre-reading skills in Grade 1 Teachers' ability to identify students at early risk for reading difficulties in Grade 1	Teachers' assessment practices Teacher ratings and students' test scores of letter knowledge and phonological awareness	Descriptive statistics Spearman's rank order correlation coefficient Binary logistic regression analysis
Sub-study II: SEN teachers' assessment practices Identification accuracy of reading fluency and compre- hension (typical/low-achiev- ing students) in Grade 6	SEN teachers' assessment practices SEN teacher ratings and students' test scores of reading fluency and reading comprehension	Descriptive statistics Spearman's rank order correlation coefficient Binary logistic regression analysis
Sub-study III: SEN teacher-reported reasons for part-time special education Students' additional teacher-reported difficulties Differences between the support groups and the control group in reading fluency, reading comprehension, and arithmetic fluency in Grades 1 and 4 Development of reading and arithmetic skills across Grades 1 to 4 in two groups of children receiving part-time special education, as compared with each other and the control group	Reasons for special education Additional learning-related difficulties Students' test scores in Grades 1, 2, 3, and 4 of read- ing fluency, reading comprehension and arithmetic fluency	Descriptive statistics Cross-tabulations χ² tests Mixed-design ANOVA

4 OVERVIEW OF THE ORIGINAL STUDIES

4.1 Sub-study I: Teachers' ability to identify children at early risk for reading difficulties in Grade 1

The aim of sub-study I was to investigate the class teachers' (n = 40) and the SEN teachers' (n = 63) assessment practices of the students' pre-reading skills and the accuracy of the teachers' assessments by comparing the teacher ratings to the students' test scores in Grade 1. The data were analyzed with the IBM SPSS program, using descriptive statistics, correlations, and binary logistic regression analyses. Pre-reading skills were tested on letter knowledge and phonological skills (phoneme identification and phoneme blending). The accuracy of the teachers' assessments was examined, based on both class teachers and SEN teachers' ratings and the students' test scores. Sensitivity of the assessments indicated the proportion of students the teachers had, according to the test scores, correctly rated as having the risk for RD (true positives). Specificity of the assessments referred to how accurately, according to the test scores, the teachers had identified those students who were developing typically in reading and did not have the risk for RD (true negatives). The identification accuracy of the assessments was counted as follows: sensitivity percent (true positives/(true positives + false negatives)* 100) and specificity percent (true negatives/(true negatives + false positives) *100).

Concerning the JLD study, most SEN teachers used only one type of assessment practice (53%), either qualitative or curriculum-based measurement (CBM). In case they used two practices (21%), they usually combined qualitative assessment (e.g., own observations, discussions) with tests. Twenty-six percent of the SEN teachers used all three practices. Almost 50% of the SEN teachers used tests, either alone or with qualitative assessment, and/or CBM. Class teachers' assessment practices of reading were not available in the JLD study. Regarding the First Steps study, class teachers used only one type of assessment practice more often (70%) as compared with the SEN teachers (37%). Approximately 60% of the class

teachers used qualitative assessment as their only practice. The most common practice used by the SEN teachers was tests, either alone or in conjunction with some other practice, covering nearly 90% of the SEN teachers. In addition, 56% of the SEN teachers relied on two types of assessment practices, most commonly tests combined with CBM. None of the class teachers and only a few (7%) SEN teachers reported using three types of assessment. Between the teacher ratings and the test scores, the associations (Spearman's correlation coefficients) were significant but moderate, except for the letter knowledge task, which correlated rather well between the SEN teacher ratings in both the JLD and the First Steps studies (.52 and .50, respectively).

Examining the accuracy of the teacher ratings against the students' test scores showed that sensitivity of the class teachers' assessments (available only in the JLD study) was low in all tested skills, unlike specificity, which reached the recommended level of accuracy in two out of three skills, in letter knowledge and phoneme identification (90% and 93%, respectively). Regarding the sensitivity of the SEN teacher ratings, there was a significant difference between the JLD and the First Steps study teachers in letter knowledge from 55% to 100%, respectively, and in phoneme blending from 72% to 100%, respectively. In particular, the specificity rates varied remarkably among the JLD and the First Steps study participants; in letter knowledge, the rates were 83% and 23% (respectively), and in phoneme blending 54% and 9% (respectively).

To conclude, the results showed that teachers used varying practices in assessing the risk for RD at school start. The accuracy of the assessments seems to depend to some extent on the assessment practice used. For instance, using relevant and appropriate tests, either alone or with some other practice, and having the necessary skills to apply the tests and interpret the results can lead to accurate assessment. Additionally, depending for the most part on qualitative assessment practices can lead to poor accuracy of identification of the struggling students in reading. To be able to reliably identify the students at risk for RD at the beginning of school, teachers need to have adequate assessment skills, proper assessment tools, and knowledge of assessment procedures.

4.2 Sub-study II: Special education teachers' identification of students' reading difficulties in Grade 6

The aim of sub-study II was to investigate the experienced relevance of assessment practices the SEN teachers (n = 29) used in evaluating their students' reading performance. Another aim was to examine how accurately the SEN teachers had been able to rate the students' reading fluency and reading comprehension (poor vs. typical performance) in comparison to the students' test scores in the spring of Grade 6. The data were analyzed with IBM SPSS program, using descriptive statistics, correlations, and binary logistic regression analyses. The iden-

tification accuracy of the assessments was counted as follows: sensitivity percentage (true positives/(true positives + false negatives) * 100) and specificity percentage (true negatives/(true negatives + false positives) * 100).

The results showed that the majority of the SEN teachers considered several assessment practices as important, which were generally used together in assessing students. Sixty-six percent of the SEN teachers reported using all five given options: own observations, discussions, tests, assessment forms and exams. The most preferred of all given options were two indirect assessment practices: own observations and discussions, which were ranked as the most important practice by 86% of the SEN teachers. Only 15% of the SEN teachers reported preferring direct assessment practices, such as achievement tests, to other given options. However, 93% of them used tests in their assessment.

Concerning reading fluency and reading comprehension, the SEN teachers' ratings were significantly but only moderately associated with the students' test scores (.39 and .24, respectively). Closer investigation revealed that the sensitivity rates of reading fluency and reading comprehension were both below the acceptable level (≥90%) and the specificity rates (≥80%; Catts et al., 2015; Compton et al., 2010). The percentage of correctly identified students with difficulties in reading fluency was 63% and in reading comprehension 70%. The percentage of correctly identified students performing typically in reading fluency was 69% and in reading comprehension 20%. This means that SEN teachers rated students having difficulties, especially in reading comprehension, in contrast with their test performance.

In conclusion, at the end of primary school, just before the transition to lower secondary school, a significant proportion of the students having RDs remained unidentified and, therefore, did not get the support they needed. Although SEN teachers are trained to identify the students' LDs, sub-study II suggests that there is still a need to improve the assessment practices and especially the tools for closer monitoring of the students' development across the primary grades after identification of the difficulty.

4.3 Sub-study III: Development of reading and arithmetic skills across Grades 1 to 4 in two groups of children receiving part-time special education

Sub-study III investigated the reasons for part-time special education, students' other learning-related difficulties reported by the SEN teachers in Grade 1, and the developmental trajectories of students' reading fluency, reading comprehension, and arithmetic fluency, based on the students' test scores, from Grade 1 to Grade 4. The students receiving part-time special education according to the SEN teachers' reports in Grade 1 (n = 130) were placed into two groups: Early-stage support group (ESS, n = 55) and Long-lasting support group (LLS, n = 75), based on the length of the support periods (1-2 years or 3-4 years, respectively). The

control group (n = 2,027) consisted of peers participating in the follow-up study but not included in the aforementioned two support groups.

First, this sub-study examined the reasons for part-time special education the SEN teachers had reported, and also the students' teacher-reported learning-related difficulties in Grade 1. Differences between the number of identified learning-related difficulties in the support groups were also investigated. The second aim of the sub-study was to examine to what extent in Grade 1 the two support groups differed from one another in reading fluency, reading comprehension, and arithmetic fluency, and whether the lengths of the support periods were reflected in the levels of those skills across Grades 1 to 4 as compared with the control group. The third aim was to demonstrate the differences in the skills among the ESS group, the LLS group, and the control group at the end of the follow-up. The analyses were conducted with the IBM SPSS program, using descriptive statistics, cross-tabulations with χ^2 tests, and mixed-design ANOVAs.

The results showed that the main reason (69%) for part-time special education among the first grade students in both support groups was RD, and the second reason (30%) was expressive language difficulty (i.e., articulation). The most commonly reported additional difficulty was attention difficulty in both groups (ESS group, 64%; LLS group, 85%). Significant differences in the number of additional difficulties between the ESS and LLS groups appeared in attention, receptive language, memory, and motivation, meaning that the students in the LLS group had those difficulties more often as compared with the other group. Socioemotional difficulty was the only additional difficulty observed more frequently by the SEN teachers among the students of the ESS group than those of the LLS group. In Grade 1, between the two support groups, there were significant differences in all tested skills, and the LLS group differed from the control group in all three skills. Between the ESS group and the control group, significant differences were found in reading fluency and reading comprehension, unlike in arithmetic fluency. At the end of Grade 4, the ESS group had reached the level of the control group in all three skills, and the only significant difference between the support groups was in reading fluency. The LLS group differed in all three skills from the control group from Grade 1 to Grade 4, and especially the development of reading fluency was slow, in comparison to the ESS group. At the end of the follow-up, in Grade 4, the differences between the LLS group and the control group remained, but no significant differences were found between the ESS group and the control group. Instead, the differences between the LLS and the ESS groups were still statistically significant in reading fluency.

To conclude, the primary school students who had persistent and multiple overlapping difficulties in learning did not catch up with their age level even though they had received long-lasting support from the SEN teachers. On the other hand, short periods of part-time special education were sufficient for students with fewer and milder LDs. Still, more frequent follow-up of the effect of the SEN teachers' support is necessary for securing the positive development of the students' academic skills.

5 GENERAL DISCUSSION

As outlined in a number of earlier studies, the role of teachers as the most relevant assessors of students' academic skills seems indisputable (Bailey & Drummond, 2006; Herppich et al., 2018; Kolovou et al., 2021). Obviously, it is important to examine how acquainted the teachers are with various assessment practices, and how appropriate assessment tools are for identification of difficulties in the students' academic skills development. The theoretical background of the present thesis lies in teachers' knowledge, skills, and practices regarding the assessment processes, identification of LDs, and the effectiveness of provided support. The context of this thesis was the Finnish primary school, where the teachers were working in mainstream settings and students receiving part-time special education were studying within general education. The thesis had two main research aims: 1) to examine the teachers' assessment practices of reading and to investigate the correspondence and the accuracy of the teachers' ratings as compared with the students' test scores, and 2) to study the students' SEN teacher-reported reasons for parttime special education and additional learning-related difficulties in Grade 1, and how the aforementioned factors and the length of part-time special education periods were reflected in the development of the students' reading fluency, reading comprehension, and arithmetic fluency skills across Grades 1 to 4. The results of this thesis demonstrate the importance of teachers' individual student assessments and the instructional decisions teachers make to enhance the development of students' academic skills. Accurate identification of the need for support and monitoring of the students' academic skills development is required for targeting the support to the students that need it the most.

5.1 Teachers' assessment practices to identify reading difficulty

To make fair and valid instructional decisions and to support their students' learning, teachers are required to master certain knowledge and skills connected with assessment, which is referred to as assessment literacy (Stiggins, 1991;

DeLuca et al., 2016). According to prior studies, teachers' assessment practices can be divided into two main categories: direct or indirect practices (Begeny et al., 2011; Woolley, 2008). The first specific aim of this thesis included the investigation of what kind of assessment practices the class teachers and the SEN teachers reported that they used in assessing the students' risk for RD at the beginning of school in Grade 1 (Sub-study I). The purpose of the teachers' assessments was to determine the students' current skills for planning the next steps in learning. The results show that the majority of the class teachers relied most on an indirect method: own observations which were used with other assessment practices or on its own. This is understandable, taking into account the small number of standardized tools available to teachers in Finland. Nevertheless, although indirect practices, used on their own, can lead to inaccurate assessments, they can give the teachers unique information about the students' skills that would not be possible to get from a standardized screening instrument (Titley et al., 2014). According to sub-study I, there were differences between the two SEN teacher samples, which can at least be partly explained by the timing of the data collections (JLD data were gathered during 2000-2003, and First Steps data in 2007) and the relatively small teacher sample sizes. Fifty-three percent of the SEN teachers participating in the JLD study reported that they used only one assessment practice, either qualitative practice or CBM. Additionally, 47% of them reported using tests, and 26% used three assessment practices. Nearly 90% of the SEN teachers that participated in the First Steps study reported using tests alone or with CBM. According to earlier evidence (Compton et al., 2010; Martin & Shapiro, 1998), using various assessment tools in parallel usually improves the accuracy of the teachers' assessments.

The present thesis also examined which assessment practices the SEN teachers preferred in the assessment of reading (Sub-study II). The results first showed that SEN teachers used a variety of assessment practices in parallel, which should have been manifested as high accuracy in the assessments. However, the results also reinforced that teachers' assessments rely, for the most part, on indirect practices, either observations or discussions with students, teacher colleagues, or parents. However, tests were considered the second most important practice by 41%, and altogether, it was the most important or the second most important practice for 56% of the SEN teachers. This finding was somewhat unexpected, as compared with the results of sub-study I, where the SEN teachers participating in both longitudinal studies reported relying mostly on qualitative practices. Nevertheless, a study by Atjonen et al. (2019) demonstrated that Finnish teachers in basic education (primary school, Grades 3 and 6) and general upper secondary education preferred summative, individual practices the most.

However, according to sub-study II, over 90% of the SEN teachers used tests in their assessments to some extent, even though they were not necessarily considered the most important practice. It seems that many teachers combined tests with qualitative assessments and not with CBMs (e.g., exams or textbook assignments), which according to prior findings can improve the accuracy of the teach-

ers' assessments (Feinberg & Shapiro, 2009). Studies have indicated that the accuracy of the assessments improves when indirect measures (e.g., observation or teacher rating scales) are used in conjunction with standardized measures (Flynn & Rahbar, 1998; Speece et al., 2011; Titley et al., 2014). As Dixson and Worrell (2018) discovered in their study, different forms of assessment can complement each other, and using them in parallel can increase the accuracy. Of course, this calls for developing a range of assessment tools for the teachers that they can choose from. Recently, for example, a digital Read Grader (RG), a screen for testing reading comprehension and technical reading across Grades 2 to 7, was successfully tested and implemented in Finland (Ståhlberg et al., 2020).

5.2 Accuracy of teachers' assessments

The first specific aim of this thesis also involved examining the accuracy of the teachers' ratings of reading compared with the students' test scores. In the present thesis, assessment accuracy was operationalized as the correlation between teacher judgments and students' test performance (see Kolovou et al., 2021). Assessment accuracy was investigated in two separate sub-studies: at the beginning of primary school (Sub-study I) and in Grade 6 (Sub-study II). The findings in sub-study I indicated that the associations between the teacher ratings and test scores were in most cases moderate. This means that the teachers' assessments of students' risk for RD had not been in line with the test performance, and therefore the support had also been targeted inaccurately. This result is in accordance with prior studies (e.g., Woolley, 2008), indicating that solely using indirect measures, such as classroom observations, can often lead to inaccurate assessments of the students' actual levels of reading. Prior studies have also found that teachers' judgment accuracy is higher in assessing typically or better achieving students than those achieving at a lower level (Bailey & Drummond, 2006; Shapiro et al., 2017; Thiede et al., 2018). It has also been found that teachers are better at judging students' overall reading ability, instead of differentiating among various subskills of reading (Paleczek et al., 2017; Südkamp et al., 2012).

The results of sub-study I are partly in accordance with prior research. In particular, class teachers were poor in identifying the first grade students having the risk for RD; however, they identified rather accurately the students without the risk for RD (see Shapiro et al., 2017). As we know that class teachers interact with their students on a daily basis, this result can be seen as surprising. The explanation could be that there is a large group of students to observe in the classroom, which can lead to an individual student's progress not being systematically monitored. The accuracy of the assessments of the SEN teachers participating in the First Steps study regarding the risk of RD in Grade 1 was excellent, since all struggling students were identified. Nevertheless, the teachers poorly recognized students who, according to the test scores, did not have a risk for RD. In letter knowledge and phoneme blending, the specificity rates were very low. One interpretation may be that, particularly at the beginning of primary school,

SEN teachers are prepared to take into account all possible signs of difficulties in learning to read. This is a way of ensuring that all children learn to read, and in Finland, even the children who are nonreaders at school start can acquire good reading accuracy during the first school year (Lerkkanen et al., 2004; Soodla et al., 2015). Another explanation may be that at the time of the data collection, at the beginning of Grade 1, the SEN teachers' views of the students' skills were based on antecedent knowledge of their skills in preschool and their need for support in that phase. Alternatively, the teachers may have trusted their own intuition (see Vanlommel & Schildkamp, 2019) and perceptions, and the decision to provide the support could have been based on insufficient information.

The rather low sensitivity rates with regard to the teacher ratings of the sixth grade students' were especially alarming. The results indicate that the assessment practices or tools used at the time were not accurate enough for about one-third of the students with poor reading test performance were not identified, which probably resulted in inadequate support for RD. Also, especially the low specificity rate with respect to the teacher ratings in reading comprehension raises a question about how accurately teachers, in general, are able to identify the students in need of support in the later phase of primary school. It has been shown that measuring students' reading comprehension is challenging because of the complexity of comprehension skills (Elleman & Oslund, 2019). In Finland, there is currently only one standardized screening test for reading comprehension in primary school (ALLU; Lindeman, 1998), and to gain accurate assessments for educational decision-making, using multiple tests or tasks (Keenan, 2016; Elleman & Oslund, 2019), together with teachers' observations and students' self-evaluations (see Peura et al., 2019, 2021) is recommended.

It has been demonstrated that, for some students, poor academic skills (e.g., poor word reading and reading comprehension) can be a learning block across compulsory education, even into adulthood (Holopainen & Hakkarainen, 2019; Landerl & Wimmer, 2008). Nevertheless, some students do not experience RD until later in primary school, when the adoption of written texts for learning academic content increases (Oslund et al., 2018). A recent study examining Finnish Grade 9 students suggests that difficulties in reading comprehension and reading fluency are two distinct difficulties, and that students with reading comprehension difficulties do not necessarily suffer from slow reading (Torppa et al., 2020). This finding shows that, even if there are no difficulties in reading fluency, they may still emerge in reading comprehension. However, the difficulty in reading is not necessarily a serious obstacle to academic achievement for all students with RD, at least in a transparent Finnish language context.

According to the findings, the correlations between the teachers' ratings and the test scores were significant but mostly moderate for students in Grades 1 and 6, which indicates the need for developing and deploying more systematic assessment tools in the identification of RD. They should also be applicable for identifying upper elementary grade students' RD as well as monitoring their learning progress. Prior studies have shown that using multiple tools (e.g., task-specific rating scales and test batteries together) in the assessment of RD can also

enhance accuracy (Kikas et al., 2017; Paleczek et al., 2017). It has been demonstrated that when teachers base their assessments mostly on qualitative practices, this can lead to inaccurate results. Nevertheless, in a study by Meissel et al. (2017), teachers had access to students' test scores in addition to their own judgments using observations and conversations, and the correlations were still average. For assessing students' skill development and making educational decisions, DBDM has proven efficient, leading to increased student achievement (Filderman et al., 2018; Hoogland et al., 2016). By using DBDM instead of casual and informal assessment practices, the reliability of the assessments can be increased, and the decisions are then based on comprehensive knowledge of the student's abilities.

In the longitudinal data used in the sub-studies, mostly group-administered tasks were used, and to discover the real nature of students' RD, the group-administered screening should be followed by individual testing for a more precise evaluation of the difficulty, when needed. In sub-studies I and II, one reason for the weak or moderate correlations between the teacher ratings and the test scores may be that teachers were asked to use a three-point rating scale without any precise cut-off points, whereas the reading tasks had continuous scales, and the cut-off point for low achievement had been distinctly specified. A known disadvantage of using rating scales is that for discrimination of students with different achievement levels or identifying the at-risk students, no clear cut-off scores usually exist (Titley et al., 2014).

5.3 Learning support and development of academic skills

The second specific aim of the thesis focused on the SEN teachers' reported reasons for part-time special education in Grade 1 and also the additional difficulties the first grade students had according to the SEN teachers. Further, the provision of part-time special education, and how its duration was reflected in the students' academic skills development across Grades 1 to 4, was investigated. A group of students (n = 130) participating in the First Steps study that had received part-time special education were followed up from Grade 1 to Grade 4. Students were divided into two support groups, based on the length of part-time special education: Early-stage support group (ESS, part-time special education in Grade 1 or in Grades 1 and 2), or Long-lasting support group (LLS, part-time special education across Grades 1 to 3 or Grades 1 to 4). The development of their reading fluency, reading comprehension, and arithmetic fluency was assessed in terms of the reasons for support, additional learning-related difficulties, and to what extent the different lengths of the support periods affected the development of the students' academic skills.

Part-time special education is a flexible means of support when a student shows signs of LDs, and sub-study III confirmed that the most frequent reason for early support at school was RD (see also Holopainen, Kiuru et al., 2017). In Finland, for instance, no formal diagnosis is needed for provision of support for students with difficulties in reading and math, and support can be provided as

long as a student needs it. For a better determination of the students who have the actual need for support in their learning, it is important that teachers are familiar with the students' strengths and weaknesses and have the possibility to monitor their progress systematically (Urhahne & Wijnia, 2021). Teachers should also make their judgment using more DBDM in their work (Mandinach et al., 2008; Pierce et al., 2013). DBDM should be a central part of teachers' assessment work in the form of systematic procedures of data collection, analysis, and interpretation, structured observation, and formal tests (Hoogland et al., 2016).

Sub-study III showed that according to SEN teachers, the most common reason for receiving part-time special education in Grade 1 was RD, and attention difficulty was reported most often as an additional difficulty. The results also demonstrated that for some students, a shorter period of support was enough for catching up to their typically achieving peers' skill levels, while for some students, part-time special education during the first four years of school seemed to be insufficient. On the group level, none of the studied skills of the students in the LLS group had developed according to expectations, and therefore the support had lasted for several years. Though their skills developed, the development was slow, and they lagged behind the ESS group and the control group in every skill, even by the time they entered Grade 4. On the other hand, the students in the ESS group developed their skills so that the gap between this group and the control group had narrowed, and in Grade 4, there were no more significant differences. SEN teachers also reported the overlapping learning-related difficulties the students had. The students in the LLS group had in general more additional difficulties than the students in the ESS group, and the differences between the groups were significant concerning attention, receptive language, memory, and motivation difficulties. Attention difficulties were the major learning-related additional difficulty in both groups, although they occurred more frequently among the students in the LLS group. These results are in line with prior studies where students demonstrating LDs that overlap with attention difficulties have shown a lack of responsiveness to support (DuPaul et al., 2013; Willcutt, 2018). The inattentive form of Attention Deficit Hyperactivity Disorder (ADHD) has been associated with RD more often than hyperactivity-impulsivity symptoms (Grills-Taquechel et al., 2013; Paloyelis et al., 2010; Prochnow et al., 2013). In substudy III, the identification of students' attention difficulties was based solely on the SEN teachers' perceptions, and no further specification of the nature of the attention difficulties (inattentive or other type) was reported. In the light of earlier studies (Breslau et al., 2010; Morgan et al., 2008), more specific identification of RD co-occurring with attention difficulties is required.

Also, students with persistent language impairment may have a higher risk for weak response to reading interventions (Denton et al., 2013). The results of this sub-study align with prior evidence that specific language impairment (SLI) and RD share common characteristics, and that the presence of language difficulty can slow a child's literacy development (Spanoudis et al., 2019). Additionally, the study supports prior findings that comorbid difficulties of RD and math

difficulty can be persistent, and those students need support for a longer period, in comparison to students with a single deficit (Koponen et al., 2018).

It has been indicated that as compared with single deficit difficulties, comorbid difficulties are usually even more resistant to early-stage intervention and support (Koponen et al., 2018). Co-occurring LDs have also been shown to slow down the development of reading and arithmetic skills across early primary grades and, in the long run, this can be associated with lower academic achievement during upper secondary education as compared with students without overlapping difficulties (Holopainen & Hakkarainen, 2019). In case a student continuously has overlapping difficulties that interfere with learning, teachers must have the means to solve a range of learning-related challenges. A situation when the student has a broader variety of learning-related difficulties overlapping with RD can lead to slow reading development (see Etmanskie et al., 2016). In a study by Ilyushina et al. (2019) the first grade students' difficulties in phonological processing were strongly associated with both reading and mathematics difficulties, which underlines the importance of early assessment of phonological skills. Prior studies have also shown the overlap of reading and attention difficulties (Grills-Taquechel et al., 2013; Kofler et al., 2019; Peterson et al., 2017) and the co-occurrence of attention difficulty with both reading and math difficulties (Child et al., 2019). Plausible reasons for these persistent, overlapping difficulties can be that, despite the need for more intensive support, the level of students' support remains constant, and it is targeted to support their several needs, which leads to less practice as compared with students with a single difficulty.

Regular monitoring of the students' progress is essential for teachers to be able to target the support directly to the core difficulty. Nevertheless, teachers may not have the time nor the competence to help struggling students (Cancio et al., 2018; Langher et al., 2017). According to sub-study III, students with comorbid and persistent LDs were provided longer support compared with the students having only one certain difficulty, for example, in reading and spelling. In a recent comparative study by Soodla et al. (2019), Finnish students outperformed their Estonian peers in reading comprehension after Grade 1. This was explained by the more effective special education support that is provided in Finland, which comprises the availability of remedial teaching, having SEN teachers in the classroom, and also the use of adaptive learning games (e.g., GraphoGame, see Richardson & Lyytinen, 2014). Nevertheless, it has been shown (see Holopainen, Kiuru et al., 2017) that even intensive, regular provision of part-time special education may not be enough for those students unless the support is carefully targeted. In Finland, according to current legislation, support for learning can be long-lasting, but it may not always be the best solution for all struggling students (Holopainen, Kiuru et al., 2017). A recent study by Solheim et al. (2020) found that the most effective means of support for slow progressing students at risk for RD was targeted early intervention, implemented in small groups. Nevertheless, for any support practice to be effective, teachers need to have adequate information about the students' current capacities and skills (Filderman et al., 2018).

5.4 Ethical considerations

This thesis was carried out following the ethical guidelines of the Finnish Advisory Board of Research Integrity (2012) and therefore meets the following criteria: 1) respecting the autonomy of research subjects, 2) avoiding harm, and 3) respecting privacy and data protection (National Advisory Broad on Research Ethics, 2009). Additionally, the JLD study was approved by the hospital district of Central Finland for the children who participated as newborns in the electroencephalography (EEG) assessments at the maternity ward. The protocol concerning the First Steps study (from which the majority of the data used in this thesis was obtained) and children below 15 years of age was submitted for ethical review by the Committee of Ethics at the University of Jyväskylä at the outset of the followup study, and a statement of approval concerning the ethics of the study was received on June 15, 2006. Participation was voluntary, and all study participants (teachers and children's caretakers) gave written consent for their own or the children's participation. Participants were also aware they were free to drop out of the study at any time. Data relating to individuals and schools were made anonymous through the allocation of code numbers. All study participants were treated equally and fairly, and the results of the study were reported respectfully. The storage of research material and confidential treatment of data were undertaken in accordance with the University's Ethics Committee Guidelines.

5.5 Practical implications

This thesis focused on a key element of primary school teachers' work, namely, assessment of the students' reading and math skills and making decisions about their need for support. Prior studies investigating Finnish teachers have mostly concentrated on class teachers or subject teachers working in regular classrooms. There are studies, for example, on teacher-student interaction (Muhonen et al., 2018; Pöysä et al., 2019) and teachers' experience of stress (Chaudhuri et al., 2021). Nevertheless, the characteristics of the SEN teachers' work, or their assessment practices, have not been widely studied thus far. In the three sub-studies of this thesis, most of the participating teachers were SEN teachers. Hopefully, the findings will extend the knowledge of teachers' assessment practices in general, especially regarding SEN teachers, and of what kind of instructional procedures guide the provision of part-time special education in Finland.

The key findings of the present thesis regarding practical implications suggest that it is crucial that teachers use multiple assessment practices to gain accurate and reliable assessments. In addition, teachers should have reliable assessment tools for individual follow-up of the students' skill development. From the inclusive education perspective, there is the need for stronger collaboration between the class teachers and the SEN teachers in the assessment of the students' risk for RD and their developing reading skills. This calls for the development of

more uniform Finnish teacher education, based on the premise of data-based decision-making.

Sub-study I indicated that at the beginning of school, the students' reading skills are being monitored in various ways by the class teachers and the SEN teachers. According to the findings, the SEN teachers used a range of assessment practices available at the time of data collection, as compared with the class teachers who mostly relied on qualitative practices. Identifying struggling students in particular was more effectively achieved among the SEN teachers. Teachers have different roles and possibilities in monitoring the development of the students' skills: the class teachers have more frequent contact with the students, while the SEN teachers can monitor the students more individually and in smaller groups as compared with the class teachers. What was a rather alarming finding was the degree of typically performing students who, according to the teacher ratings, were in need of support. The findings emphasize that teachers must have such assessment practices and tools at their disposal that can reliably enable them to distinguish the students needing support from the students performing typically.

Sub-study II showed that the SEN teachers generally used several assessment practices in parallel. Also, the current legislation in Finland (FNBE, 2014) enables teachers to choose the assessment practices, as well as the support measures, to best fit the needs of individual students. The challenge is that there is a great variety of practices that have not been systematically collected and published, and it would be worthwhile if there was a range of practices available for all teachers to choose from. The participating students were already in Grade 6, and the results reflected the importance of prevention and early support of RDs to avoid the exacerbating the difficulties. It has been shown that generally the support provided in the early phase of primary school is more effective than during later school grades (Miciak et al., 2017; Vaughn & Wanzek, 2014). Also, the support needs to be modified according to what the specific areas of students' LDs are, and what has been the response to support. These kinds of procedures are currently associated with the RTI model and DA in the United States, where instruction and assessment are merged (Grigorenko, 2009) and interact.

In sub-study III, some students' reading and math difficulties were already identified at school start, and they proved to be long-lasting, although the students had been provided with part-time special education for many years. This thesis did not examine the nature of the given support, so it remains unknown what kind of support the students had received from the SEN teachers. Even though Finnish legislation allows the provision of part-time special education whenever and as long as it seems to be needed for the students. Nevertheless, the support should be targeted more accurately to avoid providing insufficient or inefficient support to the students with LDs. As noted previously, monitoring of the student's response to provided support and modifying the support will be the major challenges of the Finnish school system in the future. To be able to accomplish that, the teachers need proper tools for systematic follow-up of the students' skill development throughout primary school.

5.6 Limitations and future directions

There are several limitations that need to be acknowledged regarding this thesis and that should be addressed in future studies. First, the children participating in the two longitudinal studies, apart from the control groups, were selected as participants based on certain characteristics, connected with the familial risk of dyslexia (JLD study), or the risk for reading difficulty determined by the researchers at the end of the kindergarten year (First Steps study). Also, neither of the follow-up studies focused specifically on the teachers' assessment knowledge, however, it was possible to examine, using the teachers' responses regarding their assessment practices on the student observation forms (sub-study I), and the SEN teachers' questionnaires (sub-study II). Especially in sub-study I the class teachers' (JLD study) response rate was low, and therefore, according to their assessment practices, the results can only be seen as indicative. However, because of the lack of studies addressing teachers' assessment practices and RD identification, the findings of the present study are important. In future research, it would be interesting to investigate the effect of teachers' training together with work experience, and how that affects the teachers' assessment knowledge and practices.

The second limitation is that the participating teachers were aware of the study aims and contents, and that the children were taking part and associated with a certain research (JLD or First Steps study), which may have had an effect on their student ratings. The students may also have been somewhat familiar to the teachers because they had received part-time special education in the earlier grades, especially concerning sub-study II. However, the teachers participating in the First Steps study were not informed which group the students in the follow-up belonged to and what their risk groups were (at-risk or control group), even though all students belonged to the group of 598 individually followed-up students.

Third, in sub-study I, the JLD data were collected a long time ago, in 2001–2003, when the students were in Grade 1 (including different age cohorts). Regarding the First Steps study, the data from Grade 1 were gathered in 2007–2008, and that of Grades 2, 3, and 4 during 2008–2011, and Grade 6 data in 2013. Thus, in respect to getting a comprehensive understanding of the teachers' assessment practices and their accuracy, using more current data would have been important. For example, the contents of teacher education concerning the participating teachers' basic assessment knowledge and available assessment tools may have had an effect on the teachers' assessment practices on which they based their student observations and ratings of the skills at the time. The data used in sub-study II are the most current, and the gathering of data was carried out in 2013, after the new learning and schooling support had come into effect in Finland. Some recent studies have investigated the implementation of the reform and its effect on teachers' work as well as teachers' attitudes toward more inclusive teaching

practices as a consequence of the reform (Lakkala & Thuneberg, 2018). Nevertheless, the longitudinal designs of both studies are valuable, and they enabled gaining comparable data from both Grades 1 and 6. In the future, the effects of the current learning and schooling support in Finland on the identification of students at risk for LD are worth studying.

The fourth limitation is associated with the teachers' possibly subjective interpretations of the given categorical classifications (clear difficulty, mild difficulty, no difficulty) concerning the students' skills (Sub-studies I and II). In particular, differentiating between mild and clear difficulties may have been challenging for the teachers because no cut-off scores were available. This limitation is also connected with the validity of the study: Did the teachers' ratings measure the same thing as the tests did? The teachers' ratings may have concerned a more comprehensive impression of the students' reading skills compared with the tests. Some teachers may have based their student ratings on a broader view of the students' reading skills, which, in turn, may have hampered the exact identification of the key difficulty (e.g., deficiency in phonological awareness).

The fifth limitation is that, according to the results, it remains rather vague which teachers' assessment tools or practices actually were the most accurate in identifying the typically performing students or the students needing support. The results indicate only what assessment practices the teachers used or preferred in their work. Additionally, with regard to sub-study III it would also have been worthwhile to examine the support practices the SEN teachers used with the students. Of course, the link between the skills assessment and support should be clear, and this calls for further research on how the assessment and the development of academic skills is related to the teacher-provided support practices. Particularly in view of the gap regarding the current knowledge of teachers' assessment practices, this should be studied more in the future.

Moreover, most of the tests used in the sub-studies were group-administered screening tests, which can also have an effect on the reliability of the results. We cannot be sure whether the students have focused properly on the tasks or understood every item correctly. Moreover, regarding reading comprehension, the study data included only one standardized test, which can also be an issue concerning reliability. In future research, to ensure comprehensive assessment data, in addition to the teachers' ratings and other available assessment practices, using both group and individual tests is recommended.

Finally, this thesis was conducted in Finland, and its education system differs in many ways from those in other countries; for instance, children enter school the year they turn seven years of age, while in the United Kingdom and U.S., the school beginners are five or six years old. The national guidelines and current legislation also allow Finnish schools and teachers relatively broad autonomy concerning how to carry out the instruction, assessment, and support (Kyttälä et al., 2021; Vitikka et al., 2012). The autonomy enables the teachers to make their own decisions using their best pedagogical expertise, but it can lead to very different procedures and practices among the municipalities or schools in regard to how the students are assessed and how their learning processes are

supported (Kyttälä et al., 2021; Pulkkinen & Jahnukainen, 2016). The findings of the present thesis highlight the importance of developing current teacher education in Finland. As a result of teacher training, all pre-service teachers should have adequate pedagogical knowledge, content knowledge, and assessment competence. The results of a recent evaluation study by Atjonen et al. (2019) suggest that to develop teachers' assessment competence during basic and continuing teacher education, assessment skills need to be a primary teacher training objective.

6 CONCLUDING REMARKS

The Finnish educational system has allocated resources for early and needsbased support for problems in learning and schooling for all students, but more effective and individually targeted means for assessment are still needed. In Finland, teachers are not obligated to use standardized follow-up procedures for instructional decisions that, however, can have long-lasting consequences for the students' educational paths. Although Finnish teacher education is considered to be of high quality, the findings of this thesis recommend that more attention should still be paid to teachers' pre- and in-service training and to continuous building of teachers' assessment competence. Also, every teacher's expertise in using various assessment practices for the early identification of students at risk for RD should be ensured. Further, the results demonstrate the need to develop standardized tools for the Finnish teachers, both for screening and individual testing (e.g., standardized tests). Nevertheless, the support resources are limited, and therefore support should be targeted to the students that need it the most. This emphasizes the importance of individual student assessment and systematic follow-up of the skill development. In sum, future studies should address teachers' assessment knowledge and effective assessment procedures, as well as the associations between the teachers' assessments and students' response to provided support.

YHTEENVETO

Tässä väitöstutkimuksessa tarkasteltiin luokanopettajien ja erityisopettajien käyttämiä lukemisen arviointimenetelmiä sekä opettajien tekemien arviointien tarkkuutta verrattuna oppilaiden testituloksiin, liittyen lukivaikeusriskin tunnistamiseen koulun alussa sekä 6. luokalla lukusujuvuuden ja luetun ymmärtämisen taitoihin. Lisäksi tutkittiin 1. luokalla alkaneen osa-aikaisen erityisopetuksen syitä ja oheisongelmia (erityisopettajien raportoimina) sekä osa-aikaisen erityisopetuksen keston yhteyttä lukusujuvuuden, luetun ymmärtämisen ja aritmetiikan sujuvuuden kehityskulkuun luokilla 1, 2, 3 ja 4. Tavoitteena oli selvittää: (1) Millaisin menetelmin luokan- ja erityisopettajat arvioivat oppilaiden lukivaikeusriskiä, lukusujuvuutta sekä luetun ymmärtämistä ja millainen yhteys opettajien arvioilla ja testituloksilla oli? Lisäksi tutkittiin, miten tarkasti opettajat pystyivät tunnistamaan ne ensimmäisen luokan oppilaat, joilla oli riski lukivaikeuksiin (osatutkimus I) tai vaikeuksia lukusujuvuudessa ja/tai luetun ymmärtämisessä 6. luokalla (osatutkimus II) sekä toisaalta ne oppilaat, jotka etenivät taitojen kehityksessä ikätasolle tyypillisesti ensimmäisellä ja kuudennella luokalla. Edelleen tutkittiin: (2) Millaisia osa-aikaisen erityisopetuksen syitä ja muita oppimisen oheisongelmia erityisopettajat raportoivat ensimmäisen luokan oppilailla, sekä erosivatko kaksi osa-aikaista erityisopetusta eri pituisen ajan saanutta oppilasryhmää ensimmäisellä ja neljännellä luokalla toisistaan ja kontrolliryhmästä lukusujuvuudessa, luetun ymmärtämisessä ja aritmetiikan sujuvuudessa sekä missä määrin erityisopetuksen kesto näyttäytyi oppilaiden edellä mainittujen taitojen kehityksessä vuosiluokilla 1, 2, 3 ja 4 (osatutkimus III).

Väitöstutkimus koostui kolmesta osatutkimuksesta, joissa käytettiin otoksia kahdesta pitkittäistutkimusaineistosta, Lapsen kielen kehitys -tutkimuksesta (Lyytinen ym., 2006) ja Alkuportaat-seurantatutkimuksesta (Lerkkanen ym., 2006-2016). Kahdessa ensimmäisessä osatutkimuksessa käytettiin poikkileikkausaineistoja 1. ja 6. luokalta ja kolmannessa oli pitkittäisasetelma luokilta 1-4. Ensimmäisessä osatutkimuksessa oli sekä LKK-tutkimuksen aineistoa (16 luokanopettajaa ja 29 erityisopettajaa sekä 33-44 oppilasta, määrä vaihteli arvioidun taidon mukaan) että Alkuportaat-tutkimuksen aineistoa (24 luokanopettajaa ja 34 erityisopettajaa sekä 69 oppilasta). Osatutkimuksissa II ja III aineisto oli kerätty Alkuportaat-seurantatutkimuksesta. Osatutkimuksessa II oli erityisopettajia 29 ja oppilaita 55, osatutkimuksessa III puolestaan erityisopettajia 35 ja yksilöseurantaan valittuja oppilaita yhteensä 130 (erityisopetusta 1. tai 1. ja 2. luokalla saaneet eli lyhyt tuki ja erityisopetusta luokilla 1-3 tai 1-4 saaneet oppilaat eli pitkä tuki) sekä kontrolliryhmä 2,027 oppilasta. Erityisopettajalla tarkoitetaan tässä tutkimuksessa osa-aikaista erityisopetusta antavaa, yleisopetuksessa toimivaa laaja-alaista erityisopettajaa (ks. esim. Takala & Ahl, 2014).

Ensimmäisessä osatutkimuksessa selvitettiin niitä arviointimenetelmiä, joita luokanopettajat ja erityisopettajat käyttivät lukivaikeusriskin tunnistamisessa ensimmäisen luokan syksyllä. Tulosten mukaan Alkuportaat-tutkimuksen

luokanopettajista 70 prosenttia käytti vain yhtä arviointimenetelmää, joko laadullista tai opetussuunnitelmaan perustuvaa menetelmää. LKK-aineistossa erityisopettajista testiarviointia käytti 47 prosenttia, kun Alkuportaat-aineiston erityisopettajilla vastaava luku oli lähes 90 prosenttia. Alkuportaat-tutkimusaineiston mukaan erityisopettajat käyttivät kahta arviointimenetelmää rinnakkain, yleisimmin testejä yhdistettynä joko laadulliseen menetelmään tai opetussuunnitelmapohjaisiin menetelmiin. LKK-tutkimuksen erityisopettajista sen sijaan noin puolet (53%) käytti kuitenkin vain yhtä arviointimenetelmää.

Lisäksi tutkittiin sitä, miten tarkasti opettajat pystyivät tunnistamaan oppilaat, joilla oli riski lukivaikeuksiin ja toisaalta sellaiset oppilaat, joilla ei riskiä ollut (sensitivity, specificity, ks. esim. Compton ym., 2010) vertaamalla opettajien tekemiä arvioita ja oppilaiden testituloksia kirjainten tunnistamisesta ja fonologisesta tietoisuudesta (alkuäänteen tunnistaminen ja/tai äänteiden yhdistäminen) toisiinsa. LKK:n osalta oli käytettävissä sekä luokanopettajien että erityisopettajien oppilasarviot, ja Alkuportaat-tutkimuksesta vain erityisopettajien tekemät oppilasarviot. Tuloksista ilmeni, että opettajien arviointien tarkkuudessa erot eri osa-alueiden välillä olivat yleisesti ottaen pieniä, mutta LKK:n luokanopettajat tunnistivat parhaiten ne oppilaat, joilla oli pulmia äänteiden yhdistämisessä, ja LKK:n sekä Alkuportaat-tutkimuksen erityisopettajat puolestaan ne oppilaat, joille kirjainten tunnistaminen oli vaikeaa. LKK:n luokanopettajat tunnistivat heikosti oppilaat, joilla oli riski lukivaikeuksiin. Sen sijaan lukutaidossa ikätasolle tyypillisesti kehittyvien oppilaiden tunnistaminen vastasi hyvin testien antamaa arviointia. Lisäksi Alkuportaat-tutkimuksen erityisopettajat tunnistivat tarkasti kaikki oppilaat, joilla oli riski lukivaikeuksiin, mutta toisaalta ikätasolle tyypillisesti lukutaidossa edistyvien oppilaiden tunnistaminen (etenkin äänteiden yhdistämisen osalta) oli heillä heikkoa. Tämä tarkoitti sitä, että tukea tarjottiin myös sellaisille oppilaille, joiden riski lukivaikeuksiin testien mukaan oli vähäinen.

Toisessa osatutkimuksessa selvitettiin, mitkä esitetyistä arviointimenetelmistä (omat havainnoinnit, keskustelut, testiarvioinnit, arviointilomakkeet, kokeet) erityisopettajien mukaan olivat tärkeimpiä lukutaidon arvioinnissa. Lisäksi tutkittiin sitä, miten tarkasti erityisopettajat olivat kuudennen luokan keväällä pystyneet arvioimaan, oliko oppilailla vaikeuksia lukusujuvuudessa ja luetun ymmärtämisessä vai suoriutuivatko he ikätasolle tyypillisesti. Opettajien tekemiä oppilasarvioita verrattiin oppilaiden testituloksiin. Tuloksista ilmeni ensinnäkin, että erityisopettajat pitivät tärkeimpinä laadullisia arviointimenetelmiä, kuten havainnointia ja keskusteluja. Testiarviointeja piti tärkeimpänä 15% opettajista, ja yhteensä 56% ilmoitti pitävänsä testejä joko tärkeimpänä tai toiseksi tärkeimpänä arviointikeinona. Kuitenkin erityisopettajista kaikkiaan 93% käytti testejä arvioinnissa jossain määrin, ja vain seitsemän prosenttia opettajista ilmoitti, että he eivät käyttäneet niitä lainkaan. Valtaosa erityisopettajista piti arvioinnissa tärkeänä useiden menetelmien käyttämistä rinnakkain. Toiseksi, erityisopettajien arviointien ja oppilaiden testitulosten väliset korrelaatiot olivat kohtalaisia sekä lukusujuvuudessa (.39) että luetun ymmärtämisessä (.24), vaikkakin hieman vahvempi yhteys oli nähtävissä lukusujuvuudessa. Sellaisten oppilaiden tunnistaminen, joilla oli vaikeuksia luetun ymmärtämisessä, oli vain vähäisessä määrin tarkempaa kuin sellaisten, joilla oli vaikeuksia lukusujuvuudessa. Toisaalta luetun ymmärtämisen osalta erityisopettajat tunnistivat heikosti oppilaita (tarkkuus 20%), jotka suoriutuivat ikätasolle tyypillisesti, eivätkä täten olisi tarvinneet tukea.

Kolmannessa osatutkimuksessa selvitettiin, millaisia syitä osa-aikaiseen erityisopetukseen erityisopettajat raportoivat ensimmäisen luokan oppilailla sekä millaisia oheisongelmia osa-aikaista erityisopetusta saavilla oppilailla opettajien mukaan ensimmäisellä luokalla oli. Alkuportaat-tutkimukseen osallistuvista, yksilöseurantaan valituista oppilaista (n = 130) muodostettiin kaksi tukiryhmää sen perusteella, miten pitkään he olivat saaneet osa-aikaista erityisopetusta. Lyhyemmän ajan eli ensimmäisellä ja/tai toisella luokalla osa-aikaista erityisopetusta saaneet oppilaat (n = 55) muodostivat yhden tukiryhmän ja toisen tukiryhmän (n = 75) muodostivat ne oppilaat, jotka olivat saaneet tukea pidemmän aikaa eli ensimmäiseltä luokalta kolmannelle tai neljännelle luokalle. Kontrolliryhmään (n = 2,027) kuuluivat ne oppilaat, jotka osallistuivat Alkuportaatseurantaan, mutta eivät olleet mukana edellä mainituissa kahdessa tukiryhmässä.

Tulokset osoittivat, että ensimmäisellä luokalla osa-aikaiseen erityisopetukseen oli monenlaisia syitä, ja joillakin oppilailla erityisopettajat raportoivat olevan useita ensisijaisia syitä. Opettajien mukaan molempien osa-aikaista erityisopetusta saaneiden tukiryhmien oppilailla (lyhyt 1-2 vuotta tai pitkä 3-4 vuotta) useimmiten syynä olivat vaikeudet lukemisessa ja/tai kirjoittamisessa (pelkkä lukivaikeus 69%) ja toiseksi eniten molemmissa tukiryhmissä raportoitiin puheen tuottamisen ongelmia (30%). Pääasiallisen syyn tai pääasiallisten syiden lisäksi opettajat raportoivat oppilailla myös muita oppimiseen liittyviä oheisongelmia. Tilastollisesti merkitseviä eroja kahden tukiryhmän oheisongelmien välillä oli tarkkaavuuden ongelmissa, puheen ymmärtämisen vaikeuksissa, muistiongelmissa ja motivaatio-ongelmissa, ja useampia vaikeuksia oli pitkään tukea saaneiden oppilaiden ryhmässä. Ainoana poikkeuksena olivat sosioemotionaaliset ongelmat, joita raportoitiin olevan jonkin verran enemmän lyhemmän aikaa tukea saaneilla oppilailla.

Kolmannessa osatutkimuksessa selvitettiin lisäksi sitä, miten kahden tukiryhmän oppilaat erosivat ensimmäisellä luokalla toisistaan lukusujuvuudessa, luetun ymmärtämisessä ja aritmetiikan sujuvuudessa sekä verrattuna kontrolliryhmään. Huomattiin, että molempien tukiryhmien välillä oli merkitsevä ero kaikissa kolmessa taidossa ja että pidempään tukea saaneiden oppilaiden ryhmä erosi myös kontrolliryhmästä jokaisessa taidossa. Sen sijaan lyhyemmän aikaa tukea saaneen ryhmän ja kontrolliryhmän välillä eroa oli lukusujuvuudessa ja luetun ymmärtämisessä, mutta aritmetiikan sujuvuudessa eroa ei ensimmäisellä luokalla ilmennyt.

Lisäksi tutkittiin sitä, millainen kehityskulku (1-4 luokilla) kahdella tukiryhmällä oli ollut testien perusteella ja kontrolliryhmään verrattuna lukemisen sujuvuudessa, luetun ymmärtämisessä ja aritmetiikan sujuvuudessa. Pidempään tukea saanut ryhmä poikkesi kaikissa kolmessa taidossa kontrolliryhmästä jokaisella luokka-asteella. Seurannan lopussa, neljännellä luokalla, erot pidempään

tukea saaneen ryhmän ja kontrolliryhmän välillä olivat säilyneet, sen sijaan merkitseviä eroja lyhyemmän aikaa tukea saaneen ryhmän ja kontrolliryhmän välillä ei enää havaittu. Neljännellä luokalla pidemmän aikaa tukea saaneen ryhmän oppilaat eivät olleet tavoittaneet lukusujuvuudessa toista tukiryhmää, vaan ero oli edelleen merkitsevä. Tuloksista voidaan päätellä, että mikäli oppilaalla on yhtä aikaa useita oppimisen vaikeuksia, niin silloin arviointi on opettajille haastavaa ja tuen kohdentaminen vaikeutuu, mikä puolestaan heikentää tuen vaikuttavuutta. Tämä tulos on linjassa aiempien tutkimusten kanssa, jotka osoittavat, että oppimisen vaikeuksilla on taipumuksena kasaantua, jolloin samalla oppilaalla on useita, päällekkäistyneitä, oppimista haittaavia ongelmia (comorbidity, Cirino et al., 2018).

Väitöstutkimuksen tulokset vahvistavat aiempien tutkimusten tuloksia siitä, että opettajilla on hyvät mahdollisuudet ja myös keinoja tunnistaa oppilaiden oppimisen pulmia sekä arvioida tuen tarvetta. Opettajien arviointiosaaminen ja opettajien tekemä arviointi on myös edellytys toimivalle ja oikein kohdennetulle oppimisen tuelle. Erityisopetuksen resurssit ovat kuitenkin rajalliset, ja sen vuoksi olisi kiinnitettävä enemmän huomiota siihen, ketkä oppilaat hyötyisivät eniten erityisopettajan tuesta. Tutkimustulosten perusteella vaikuttaa siltä, että joillekin oppilaille tukea annetaan ikään kuin varulta, vaikka he eivät sitä testitulosten perusteella tarvitsisi. Tähän saattaa olla selityksenä tuen vaikuttavuuden seurannan tai seurantavälineiden puute, jolloin opettajan arvio oppilaan taidoista ei perustu tosiasialliseen tietoon oppilaan senhetkisestä taitotasosta. Opettajan arvioinnin tulisi pohjautua riittävän usein toistettujen testien tuloksiin, jotka mittaisivat tiettyjä selkeästi rajattuja lukemisen ja matematiikan taitojen osa-alueita. Saattaa myös olla mahdollista, että etenkin myöhemmässä vaiheessa peruskoulua opettajan arvioon voi vaikuttaa oppilaan alemmilla luokilla saama erityisopetus ja oletus siitä, että tuen tarve olisi edelleen olemassa.

Opettajien olisi ensiarvoisen tärkeää tunnistaa kaikki lukemisessaan tukea tarvitsevat oppilaat mahdollisimman varhain ja tarjota heille tukea, jotta vaikeudet eivät syvenisi ja pitkittyisi. Tämän väitöstutkimuksen tulokset osoittavat, että myös kuudesluokkalaisten lukusujuvuuden ja luetun ymmärtämisen seurantaan olisi tärkeää kiinnittää erityistä huomiota, koska siirtymävaiheessa yläkouluun luettavien tekstien määrä lisääntyy ja oppimisen asiasisällöt vaikeutuvat. Yksilöllisesti räätälöidyn ja kohdennetun tuen tarjoaminen on keskeistä, mutta myös tuen vaikutusta taidon kehittymiseen tulisi säännöllisesti arvioida. Tämä olisi erityisopetuksen resurssien käytön ja suuntaamisen kannalta järkevää, niin että tukea tarvitsemattomat oppilaat eivät myöskään peruskoulun myöhäisemmässä vaiheessa veisi turhaan resursseja tukea eniten tarvitsevilta oppilailta. Olisikin tärkeää kehittää yksilöllisen kehityksen seurantaan soveltuvia menetelmiä, joilla peruskoulun aikana tietyin kriteerein ja säännöllisin määräajoin voitaisiin tarkistaa tuen oikea kohdentuminen ja luotettavasti osoittaa tarjotun tuen vaikutus kyseisen taidon kehittymiseen. Täten tukea voisi myös perustellusti muokata oppilaan tarpeiden mukaan ja tarvittaessa lopettaa. Vallitsevan lainsäädännön mukainen käytäntö, jossa oppilas saa tukea niin kauan kuin hänen katsotaan sitä tarvitsevan, ei vielä näyttäisi mahdollistavan riittävän yksilöllistä, systemaattista oppilaan taitojen kehityksen seurantaa.

Väitöksen kolmas osatutkimus osoitti, että kun tukea saaneiden oppilaiden taidot (lyhyemmän aikaa tukea saanut ryhmä) olivat kehittyneet iän mukaiselle tasolle, niin osa-aikainen erityisopetus voitiin tukimuotona lopettaa. Näin erityisopetusresursseja pystyttiin osoittamaan enemmän oppilaille, joiden vaikeudet olivat sitkeämpiä tai haasteellisempia tukea. Opettajien tekemän arvioinnin tarkkuutta voisi parantaa, jos arviointivälineissä olisi enemmän valinnanvaraa, koska mahdollisimman luotettavan arvioinnin saavuttamiseksi erityyppisiä välineitä (esim. testejä ja havainnointia) on tärkeää käyttää rinnakkain. Lukemisen arviointiin onkin viime vuosina Suomessa kehitetty erityisesti digitaalisia ja pelillisiä lukemisen arviointivälineitä (esim. Lukuseula, DigiLukiseula nuorille ja aikuisille) ja niistä on saatu jo hyviä kokemuksia. ALLU-testipatteriston (Lindeman, 1998) luetun ymmärtämisen testi on saatavilla päivitettynä myös digitaalisessa muodossa, ja lisäksi on olemassa yläkouluikäisten lukivaikeuksien seulontamenetelmä. Opettajien tekemän arvioinnin ja päätöksenteon olisi tärkeää olla systemaattista ja dataan perustuvaa (data-based desicion making, DBDM, Filderman ym., 2018), jolloin ne pohjautuvat monipuoliseen, eri menetelmin saatuun tietoon oppilaan senhetkisestä osaamisesta.

Tässä väitöstutkimuksessa arviointia ei tutkittu siitä näkökulmasta, että millaiset yksittäiset arviointimenetelmät tai arviointivälineet ovat tarkimpia, vaan tulokset jäivät yleisemmälle tasolle (osatutkimukset I ja II). Jatkossa olisikin tärkeää saada vielä yksityiskohtaisempaa tutkimustietoa siitä, millaiset opettajien käyttämät arviointivälineet ja -menetelmät tuottavat luotettavimman arvion oppilaan taidoista. Osa-aikaisen erityisopetuksen kestoon ja taitojen kehittymiseen liittyen (osatutkimus III) olisi ollut mielenkiintoista selvittää myös sitä, millaisia tukitoimia erityisopettajat olivat eri luokilla käyttäneet ja miten tuen vaikuttavuutta oli arvioitu. Jatkotutkimuksessa nämä asiat olisi tärkeää huomioida, niin saataisiin kattavampi kuva opettajien arvioinnin ja tarjottujen tukitoimien yhteydestä oppilaiden akateemisten taitojen kehittymiseen. Myös opettajien perus- ja täydennyskoulutuksessa olisi tärkeää huomioida laaja-alaisemmin opettajien arviointiosaamisen (OpAO, Atjonen, 2021) rooli opettajan työssä, etenkin kun oppimisen ja koulunkäynnin tuen myötä erityisopettajien lisäksi myös luokanopettajilta vaaditaan entistä enemmän "diagnostista silmää" oppilaiden tuen tarpeen tunnistamiseen. Ei pidä myöskään unohtaa luokanopettajien, aineenopettajien ja erityisopettajien välistä yhteistyötä, jota on mahdollista vahvistaa entistä yhdenmukaisemman opettajakoulutuksen kautta, esimerkiksi lukivaikeuksien tunnistamiseen ja arviointiin liittyen. Tämän tutkimuksen tulokset voivat hyödyttää maassamme eri koulutusasteilla toimivia opettajia ja sen lisäksi etenkin niitä maita, joissa on käytössä vastaavan tyyppinen oppimisen tuen järjestelmä ja osa-aikaisen erityisopetuksen toimintamalli.

REFERENCES

- Adlof, S. M., & Hogan, T. P. (2018). Understanding dyslexia in the context of developmental language disorders. *Language, Speech & Hearing Services in Schools*, 49(4), 762–773. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0049
- Al Otaiba, S., Baker, K., Lan, P., Allor, J., Rivas, B., Yovanoff, P., & Kamata, A. (2019). Elementary teacher's knowledge of response to intervention implementation: A preliminary factor analysis. *Annals of Dyslexia*, 69(1), 34–53. https://doi.org/10.1007/s11881-018-00171-5
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association Publishing.
- Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x
- Aro, M. (2006). Learning to read: The effect of orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531–550). Lawrence Erlbaum Associates.
- Aro, M. (2017). Learning to read Finnish. In L. Verhoeven & C. Perfetti (Eds.), Learning to read across languages and writing systems (pp. 416–436). Cambridge University Press. https://doi.org/10.1017/9781316155752.017
- 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/10.017.S0142716403000316
- Artelt, C., & Gräsel, C. (2009). Diagnostische Kompetenz von Lehrkräften. [Diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie*, 23(3), 157–160. https://doi.org/10.1024/1010-0652.23.34.157
- Atjonen, P. (2021). Opettajien arviointiosaamisen tietoperusta, arviointikäsitykset ja arviointikäytänteitä haastavat kompromissit. [Teacher assessment literacy: knowledge base, assessment conceptions, and compromises challenging assessment practices]. NMI-Bulletin, 31(2), 4–19. Retrieved January 9, 2022, from https://bulletin.nmi.fi/wp-content/uploads/2021/06/Bulletin_2_2021_E_atjonen.pdf
- Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal*, 25(2), 238–259. https://doi.org/10.1080/09585176.2013.874952
- Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kamppi, P., Rumpu, N., Hietala, R., & Immonen, J. (2019). "So that we know where we stand" Assessment of learning and competence in basic education and general upper secondary education. Finnish Education Evaluation Centre (FINEEC).
- Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004).

 Developmental dynamics of math performance from preschool to Grade 2. *Journal of Educational Psychology*, 96(4), 699–713.

 https://doi.org/10.1037/0022-0663.96.4.699

- Aunola, K., & Räsänen, P. (2007). *Laskutaito testit* [Numeracy tests]. Unpublished test material. University of Jyväskylä.
- Auphan, P., Ecalle, J., & Magnan, A. (2019). Computer-based assessment of reading ability and subtypes of readers with reading comprehension difficulties: A study in French children from G2 to G9. *European Journal of Psychology of Education*, 34(3), 641–663. https://doi.org/10.1007/s10212-018-0396-7
- Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11, 149–178. https://doi.org/10.1207/s15326977ea1103&4_2
- Ball, C. R., & O'Connor, E. (2016). Predictive utility and classification accuracy of oral reading fluency and the measures of academic progress for the Wisconsin Knowledge and Concepts Exam. *Assessment for Effective Intervention*, 41(4), 195–208. https://doi.org/10.1177/1534508415620107
- Bayat, K., & Rezaei, A. (2015). Importance of teachers' assessment literacy. *International Journal of English Language Education*, *3*(1), 139–146. https://doi.org/10.5296/ijele.v3i1.6887
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. https://doi.org/10.1037/1045-3830.23.1.43
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, 40(1), 23–38. https://doi.org/10.1080/02796015.2011.12087726
- Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, 39, 370–407. https://doi.org/10.3102/0091732X14554179
- Bernfeld, L. E. S., Morrison, T. G., Sudweeks, R. R., & Wilcox, B. (2013). Examining the reliability of reading comprehension ratings of fifth grade students' oral retellings. *Literacy Research and Instruction*, 52(1), 65–86. https://doi.org/10.1080/19388071.2012.702189
- Bigozzi, L., Tarchi, C., Pinto, G., & Gamannossi, B. A. (2015). Predicting dyslexia in a transparent orthography from Grade 1 literacy skills: A prospective cohort study. *Reading and Writing Quarterly*, 32(4), 951–977. https://doi.org/10.1080/10573569.2014.988310
- Björn, P., Aro, M., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learning Disability Quarterly*, 39(1), 58–66. https://doi.org/10.1177/0731948715594787
- Björn, P., Aro, M., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2018). Response-to-intervention in Finland and the United States: Mathematics learning

- support as an example. *Frontiers in Psychology, 9,* 800. https://doi.org/10.3389/fpsyg.2018.00800
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–144. https://doi.org/10.1177/003172171009200119
- Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2012). 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
- Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46(1), 3–18. https://doi.org/10.1111/j.1469-7610.2005.00374.x
- Campbell, M. B. (2001). Inquiry into reading assessment: Teachers' perceptions of effective practices. *Reading Horizons, A Journal of Literacy and Language Arts, 42*(1). Retrieved January 22, 2022, from https://scholarworks.wmich.edu/reading_horizons/vol42/iss1/10
- Cancio, E. J., Larsen, R., Mathur, S. R., Estes, M. B., Johns, B., & Chang, M. (2018). Special education teacher stress: Coping strategies. *Education and Treatment of Children*, 41(4), 457–481. https://doi.org/10.1353/etc.2018.0025
- 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, 166–181. https://doi.org/10.1037/a0025323
- Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2009). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*, 331–361. https://doi.org/10.1207/s1532799xssr0304_2
- Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, 48(3), 281–297. https://doi.org/10.1177/0022219413498115
- Centre for Learning Research. (2021). University of Turku.

 https://www.utu.fi/en/university/faculty-of-social-sciences/centre-for-learning-research
- Chaudhuri, S., Muhonen, H., Pakarinen, E., & Lerkkanen, M.-K. (2021). Teachers' focus of attention in first-grade classrooms: Exploring teachers experiencing less and more stress using mobile eye-tracking. *Scandinavian Journal of Educational Research*. https://doi.org/10.1080/00313831.2021.1958374
- Cho, E., Compton, D. L., & Josol, C. K. (2020). Dynamic assessment as a screening tool for early identification of reading disabilities: A latent change score approach. *Reading and Writing*, 33(3), 719–739. https://doi.org/10.1007/s11145-019-09984-1
- Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002

- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141–146. https://doi.org/10.1037/0022-0663.78.2.141
- Collins, A. C., Lindström, E. R., & Compton, D. L. (2017). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities*, *51*(2), 108–123. https://doi.org/10.1177/0022219417704636
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327–340. https://doi.org/10.1037/a0018448
- Dandache, S., Wouters, J., & Ghesquière, P. (2014). Development of reading and phonological skills of children at family risk for dyslexia: A longitudinal analysis from kindergarten to sixth grade. *Dyslexia* 20(4), 305–329. https://doi.org/10.1002/dys.1482
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, 17(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6
- Deno, S. L. (1985). Curriculum-Based Measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232. https://doi.org/10.1177/001440298505200303
- Deno, S. L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education*, *37*(3), 184–192. https://doi.org/10.1177/00224669030370030801
- Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading*, 15(2), 109–135. https://doi.org/10.1080/108884310031003623546
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, *55*, 153–159. https://doi.org/10.1080/00405841.2016.1148989
- Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In J. Dolin & R. Evans (Eds.), *Transforming assessment*. *Through an interplay between practice, research and policy*. (pp. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3

- Duff, F. J., & Clarke, P. J. (2011). Practitioner review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *The Journal of Child Psychology and Psychiatry*, 52(1), 3–12. https://doi.org/10.1111/j.1469-7610.2010.02310.x
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produced underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction*, 22, 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003
- DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications of DSM-5 for assessment and treatment. *Journal of Learning Disabilities*, 46(1), 43–51. https://doi.org/10.1177/0022219412464351
- Ehri, L. C. (1994). Development of the ability to read words: Update. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 323–358). International Reading Association.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., & Shanahan, T. (2001). Phonemic awareness instruction helps children to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250–287. https://doi.org/10.1598/RRQ.36.3.2
- 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. https://doi.org/10.1037/a0037121
- 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
- Elleman, A. M., & Oslund, E. L. (2019). Reading comprehension research: Implications for practice and policy. *Policy Insights from the Behavioral and Brain Sciences*, 6(1), 3–11. https://doi.org/10.1177/2372732218816339
- Eloranta, A.-K., Närhi, V., Eklund, K., Ahonen, T., & Aro, T. (2019). Resolving reading disability: Childhood predictors and adult-age outcomes. *Dyslexia*, 25(1), 20–37. https://doi.org/10.1002/dys.1605
- 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
- European Commission. (2013). *Eurostat Regional Yearbook* 2013: *Education*. http://ec.europa.eu/eurostat/en/web/products-statistical-books/-/KS-HA-13-001-04
- Fedora, P. (2015). Teacher judgment of early reading difficulty. *The Educational Forum*, 79(3), 280–292. https://doi.org/10.1080/00131725.2015.1037507
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*, 52–65. https://doi.org/10.1521/scpq.18.1.52.20876

- Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research*, 102(6), 453–462. https://doi.org/10.3200/JOER.102.6.453-462
- Filderman, M. J., Toste, J. R., Didion, L. A., Peng, P., & Clemens, N. H. (2018). Data-based decision making in reading interventions: A synthesis and meta-analysis of the effects for struggling readers. *The Journal of Special Education*, 52(3), 174–187. https://doi.org/10.1177/0022466918790001
- Finnish Advisory Board on Research Integrity. (2012). *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. https://tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf
- Finnish Government. (1998). Law 628. *Perusopetuslaki*. [Basic Education Act]. http://www.finlex.fi/fi/laki/alkup/1998/19980628
- Finnish Government. (2010). Law 642. *Laki perusopetuslain muuttamisesta*. [Act on Amendment of the Basic Education Act]. http://www.finlex.fi/fi/laki/alkup/2010/20100642
- Finnish National Board of Education. (2004). *Perusopetuksen opetussuunnitelman perusteet*. [Core curriculum for basic education]. https://www.oph.fi/fi/koulutus-ja-tutkinnot/perusopetuksen-opetussuunnitelman-perusteet
- Finnish National Board of Education. (2014). *Perusopetuksen opetussuunnitelman perusteet*. [Core curriculum for basic education]. https://www.oph.fi/fi/koulutus-ja-tutkinnot/perusopetuksen-opetussuunnitelman-perusteet
- Finnish National Board of Education. (2010). *Perusopetuksen opetussuunnitelman perusteiden muutokset ja täydennykset 2010* [Amendments and additions to the national core curriculum for basic education 2010].

 http://www.oph.fi/download/
 /132882_Perusopetuksen_opetussuunnitel-man_perusteiden_muutokset_ja_taydennykset2010.pdf">ja_taydennykset2010.pdf
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, J. A. (2007). *Learning disabilities: From identification to intervention*. Guilford Press.
- Fletcher, J. M., & Satz, P. (1984). Test-based versus teacher-based predictions of academic achievement. A three-year longitudinal follow-up. *Journal of Pediatric Psychology*, 9(2), 193–201. https://doi.org/10.1093/jpepsy/9.2.103
- Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools*, *35*(2), 163–172. <a href="https://doi.org/10.1002/(SICI)1520-6807(199804)35:2<163::AID-PITS8>3.0.CO;2-Q">https://doi.org/10.1002/(SICI)1520-6807(199804)35:2<163::AID-PITS8>3.0.CO;2-Q
- Fuchs, D., Fuchs, L. S., & Compton, D. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*, 78, 263–279. https://doi.org/10.1177/001440291207800301
- Gaab, N., Turesky, T. K., & Sanfilippo, J. (2019). Early identification of children at risk for reading difficulty. Neurobiology, screening and evidence-based response, and educational technology. In J. A. Washington, D. L.

- Compton, & P. McCardle (Eds.), *Dyslexia: Revisiting etiology, diagnosis, treatment, and policy* (pp. 44–56). Brookes Publishing.
- García, J. R. & Cain, K. (2014). Decoding and reading comprehension. *Review of Educational Research*, 84(1), 74–111. https://doi.org/10.3102/0034654313499616
- Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities*, 45(4), 291–307. https://doi.org/10.1177/0022219410392046
- Georgiou, G. K., Torppa, M., Landerl, K., Desrochers, A., Manolitsis, G., de Jong, P. F., & Parrila, R. (2019). Reading and spelling development across languages varying in orthographic consistency: Do their paths cross? *Child Development*, 91(2), e266–e279. https://doi.org/10.1111/cdev.13218
- Glogger-Frey, I., Herppich, S., & Seidel, T. (2018). Linking teachers' professional knowledge and teachers' actions: Judgment processes, judgments and training. *Teaching and Teacher Education*, 76, 176–180. https://doi.org/10.1016/j.tate.2018.08.005
- Gough, S. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.
- Government of Finland. (2010). *Perusopetuslaki 628/1998*. [Basic Education Act 628/1998. Amendments up to 1136/2010]. http://www.finlex.fi/en/laki/kaannokset/1998/en19980628.pdf
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42, 111–132. https://doi.org/10.1177/0022219408326207
- Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., Denton, C. A., & Taylor, P. (2013). Anxiety and inattention as predictors of achievement in early elementary school children. *Anxiety, Stress & Coping*, 26(4), 391–410. https://doi.org/10.1080/10615806.2012.691969
- Grimm, R. P., Solari, E. J., McIntyre, N. S., & Denton, C. A. (2018). Early reading skills profiles in typically developing and at-risk first grade readers to inform targeted early reading instruction. *Journal of School Psychology*, 69, 111–126. https://doi.org/10.1016/j.jsp.2018.05.009
- Halinen, I., & Järvinen, R. (2008). Towards inclusive education: The case of Finland. *Prospects*, *38*(1), 77–97. https://doi.org/10.1007/s11125-008-9061-2
- Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J.-M., & Richardson, U. (2020). Identification of reading difficulties by digital game-based assessment technology. *Journal of Educational Computing Research*, *58*(5), 1003–1028. https://doi.org/10.1177/0735633120905309
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140–145. https://doi.org/10.1177/003172170708900210
- Herppich, S., Praetorius, A.-K., Foerster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Boehmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A.,

- Boehmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. https://doi.org/10.1016/j.tate.2017.12.001
- Hintikka, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency: Is it important to practice reading aloud and is generalization possible? *Annals of Dyslexia*, *58*, 59–79. https://doi.org/10.1007/s11881-008-0012-7
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. https://doi.org/10.3102/00346543059003297
- Holopainen, L., Ahonen, T., & Lyytinen, H. (2001). Predicting delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities*, 34(5), 401–414. https://doi.org/10.1080/00313830120115624
- Holopainen, L., & Hakkarainen, A. (2019). Longitudinal effects of reading and/or mathematical difficulties: The role of special education in graduation from upper secondary education. *Journal of Learning Disabilities* 52(6), 456–467. https://doi.org/10.1177/0022219419865485
- Holopainen, L., Kiuru, N., Mäkihonko, M., & Lerkkanen, M.-K. (2017). The role of part-time special education supporting students with reading and spelling difficulties from grade 1 to grade 2 in Finland. *European Journal of Special Needs Education*, 33(3), 316–333. https://doi.org/10.1080/08856257.2017.1312798
- Holopainen, L., Taipale, A., & Savolainen, H. (2017). Implications of overlapping difficulties in mathematics and reading on self-concept and academic achievement. *International Journal of Disability, Development and Education*, 64(1), 88–103. https://doi.org/10.1080/1034912X.2016.1181257
- Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, 60, 377–386. https://doi.org/10.1016/j.tate.2016.07.012
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading and Writing Quarterly*, 25, 4–32. https://doi.org/10.1080.10573560802491208
- Ilyushina, N., Kuzmina, Y., & Kaiky, D. (2019). The deficit of phonological processing associated with both maths and reading difficulties rather than separate maths or reading difficulties. *SSRN Electronic Journal*, January 2019. https://doi.org/10.2139/ssrn.3404125
- Jahnukainen, M., & Itkonen, T. (2016). Tiered intervention: History and trends in Finland and the United States. *European Journal of Special Needs Education*, 31(1), 140–150. https://doi.org/10.1080/08856257.2015.1108042
- Jahnukainen, M., & Itkonen, T. (2018). Steps to inclusion? The role of tiered intervention in Finland and in the United States. In A. Köpfer, J. W. Justin,

- & R. Zahnd (Eds.), *International handbook of inclusive education* (pp. 345–356). Verlag Barbara Budrich.
- Jordan, J.-A., Mulhern, G., & Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *Journal of Experimental Child Psychology*, 103, 455–468. https://doi.org/10.1016/j.jecp.2009.01.011
- Karst, K., Dotzel, S., & Dickhäuser, O. (2018). Comparing global judgments and specific judgments of teachers about students' knowledge: Is the whole sum of its parts? *Teaching and Teacher Education*, 76, 194–203. https://doi.org/10.1016/j.tate.2018.01.013
- Karst, K., Schoreit, E., & Lipowsky, F. (2014). Diagnostische Kompetenzen von Mathematiklehrern und ihr Vorhersagewert für die Lernentwicklung von Grundschulkindern [Diagnostic competencies of math teachers and their impact on the learning development of elementary school children]. Zeitschrift für Pädagogische Psychologie, 28(4), 237–248. https://doi.org/10.1024/1010-0652/a00013
- Kaufmann, E. (2020). How accurately do teachers judge students? Re-analysis of Hoge and Coladarci (1989) meta-analysis. *Contemporary Educational Psychology*, 63, 101902. https://doi.org/10.1016/j.cedpsych.2020.101902
- Keenan, J. M. (2016). Assessing the assessments: Reading comprehension tests. *Perspectives on Language and Literacy*, 42(2), 17–21.
- Kent, S. C., Wanzek, J., & Yun, J. (2019). Screening in the upper elementary grades: Identifying fourth-grade students at-risk for failing the state reading assessment. *Assessment for Effective Intervention*, 44(3), 160–172. https://doi.org/10.1177/1534508418758371
- Kikas, E., Silinskas, G., & Soodla, P. (2015). The effects of children's reading skills and interest on teacher perceptions of children's skills and individualized support. *International Journal of Behavioral Development*, 39(5), 402–412. https://doi.org/10.1177/0165025415573641
- Kikas, E., Soodla, P., & Mägi, K. (2017). Teacher judgments of student reading skills: Associations with child- and classroom related factors. *Scandinavian Journal of Educational Research*, 62(5), 783–797. https://doi.org/10.1080/00313831.2017.1307271
- Kim, Y.-S., Petscher, Y., & Foorman, B. (2015). The unique relation of silent reading fluency to end-of-year reading comprehension: Understanding individual differences at the student, classroom, school, and district levels. *Reading and Writing*, 28, 131–150. https://doi.org/10.1007/s11145-013-9455-2
- Kingsdorf, S., & Krawec, J. (2014). Error analysis of mathematics word problem solving across students with and without learning disabilities. *Learning Disabilities Research and Practice*, 29, 66–74. https://doi.org/10.1111/ldrp.12029
- Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.). *Children's reading comprehension and assessment* (pp. 3–12). Routledge.

- 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
- Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*, 456–460. https://doi.org/10.1518/001872008X28385
- Kofler, M. J., Spiegel, J. A., Soto, E. F., Irwin, L. N., Wells, E. L., & Austin, K. E. (2019). Do working memory deficits underlie reading problems in Attention-Deficit/Hyperactivity Disorder (ADHD)? *Journal of Abnormal Child Psychology*, 47(3), 433–446. https://doi.org/10.1007/s10802-018-0447-1
- Kolovou, D., Naumann, A., Hochweber, J., & Praetorius, A.-K. (2021). Content-specificity of teachers' judgment accuracy regarding students' academic achievement. *Teaching and Teacher Education*, 100. https://doi.org/10.1016/j.tate.2021.103298
- Koponen, T., Aro, M., Poikkeus, A.-M., Niemi, P., Lerkkanen, M.-K., Ahonen, T., & Nurmi, J.-E. (2018). Comorbid fluency difficulties in reading and math: Longitudinal stability across early grades. *Exceptional Children*, 84(3), 298–311. https://doi.org/10.1177/0014402918756269
- Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency:
 Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45(2), 230–251. https://doi.org/10.1598/rrq.45.2.4
- Kyttälä, M., Björn, P. M., Rantamäki, M., Närhi, V., & Aro, M. (2021). Assessment conception patterns of Finnish pre-service special needs teachers: The contribution of prior studies and teaching experience. *European Journal of Special Needs Education*. https://doi.org/10.1080/08856257.2020.1853972
- Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry*, 62, 896–902. https://doi.org/10.1001/archpsyc.62.8.896
- Lakkala, S., & Thuneberg, H. (2018). Teachers' perceptions of educational reform aimed at inclusion. *Education in the North*, 25(1–2), 295–319. https://doi.org/10.26203/m93w-hw77
- Landerl, K., Freudenthaler, H. H., Heene, M., De Jong, P., Desrochers, A., Manolitsis, G., Parrila, R., & 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, 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
- Langher, V., Caputo, A., & Ricci, M. A. (2017). The potential role of perceived support for reduction of special education teachers' burnout. *International Journal of Educational Psychology*, 6(2), 120–147. https://doi.org/10.17583/ijep.2017.2126
- 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
- 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. https://doi.org/10.1023/A:1011117620895
- Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J.-E. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. *Learning and Instruction*, *18*, 548–564. https://doi.org/10.1016/j.learninstruc.2007.11.004
- Lerkkanen, M.-K. (2007). The beginning phases of reading literacy instruction in Finland. In P. Linnakylä & I. Arffman (Eds.), *Finnish reading literacy. When quality and equity meet.* (pp. 155–174). University of Jyväskylä, Institute of Educational Research.
- Lerkkanen, M.-K., Kikas, E., Pakarinen, E., Poikonen, P.-L., & Nurmi, J.-E. (2013). Mothers' trust toward teachers in relation to teaching practices. *Early Childhood Research Quarterly*, 28(1), 153–165. https://doi.org/10.1016/j.ecresq.2012.04.005
- Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006–2016). *Alkuportaat [The First Steps study]*. Finland: University of Jyväskylä, University of Turku, and University of Eastern Finland. https://www.jyu.fi/alkuportaat
- Lerkkanen, M.-K., & Poikkeus, A.-M. (2009). *Reading comprehension: Sentences*. Unpublished measurement. University of Jyväskylä.
- Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004). The developmental dynamics of literacy skills during the first grade. *Educational Psychology*, 24(6), 793–810. https://doi.org/10.1080/0144341042000271782
- Lindeman, J. (1998). *ALLU Ala-asteen lukutesti [ALLU Reading Test for Primary School]*. University of Turku, Finland: The Centre for Learning Research.
- Locascio, G., Mahone, M., Eason, S. H., & Cutting, L. E. (2010). Executive dysfunction among children with reading comprehension deficits. *Journal of Learning Disabilities*, 43(5), 441–454. https://doi.org/10.1177/0022219409355476
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018).

 Reconceptualizing the role of teachers as assessors: Teacher assessment

- identity. Assessment in Education: Principles, Policy & Practice, 25(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090
- Lyon, E. G. (2013). Conceptualizing and exemplifying science teachers' assessment expertise. *International Journal of Science Education*, 53(7), 1208–1229. https://doi.org/10.1080/09500693.2013.770180
- 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., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly*, 52(3), 514–546. https://doi.org/10.1353/mpq.2006.0031
- Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, 102(3), 701–711. https://doi.org/10.1037/a0019135
- Marchand, G. C., & Furrer, C. J. (2014). Formative, informative, and summative assessment: The relationship among curriculum-based measurement of reading, classroom engagement, and reading performance. *Psychology in the Schools*, *51*(7), 659–676. https://doi.org/10.1002/pits.21799
- Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools*, 48(4), 343–356. https://doi.org/10.1002/pits.20558
- Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4-6 years of age. *Journal of Abnormal Child Psychology*, 36(3), 399–410. https://doi.org/10.1007/s10802-007-9186-4
- McKevett, N. M., & Kiss, A. J. (2019). The influence of data on teachers' judgments of students' early reading and math skills. *Psychology in the Schools*, 56(7), 1157–1172. https://doi.org/10.1002/pits.22256
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x
- Meisinger, E., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of Dyslexia*, 60(1), 1–17. https://doi.org/10.1007/s11881-009-0031-z
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. https://doi.org/10.1016/j.tate.2017.02.021
- Miciak, J., Roberts, G., Taylor, W. P., Solis, M., Ahmed, Y., Vaughn, S., & Fletcher, J. M. (2017). The Effects of one versus two years of intensive reading intervention implemented with late elementary struggling

- readers. Learning *Disabilities Research & Practice*, 33(1), 24–36. https://doi.org/10.1111/ldrp.12159
- Moll, K., Landerl, K., Snowling, M. J., & Schulte-Körne, G. (2019). Understanding comorbidity of learning disorders: Task-dependent estimates of prevalence. *The Journal of Child Psychology and Psychiatry*, 60(3), 286–294. https://doi.org/10.1111/jcpp.12965
- Mononen, R., & Aunio, P. (2013). Early mathematical performance in Finnish kindergarten and Grade one. *LUMAT International Journal on Math Science and Technology Education*, 1(3), 245–261. https://doi.org/10.31129/lumat.v1i3.1104
- Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, 41(5), 417–436. https://doi.org/10.1177/0022219408321123
- Muhonen, H., Pakarinen, E., Poikkeus, A.-M., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2018). Quality of educational dialogue and association with the students' academic performance. *Learning and Instruction*, *55*, 67–79. https://doi.org/10.1016/j.learninstruc.2017.09.007
- National Advisory Board on Research Ethics. (2009). *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf
- Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Leaning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773
- Official Statistics of Finland. (2019). *Special education* [e-publication]. Statistics Finland. http://www.stat.fi/til/erop/tau_en.html
- Ohle, A., & McElvany, N. (2015). Teachers' diagnostic competence and their practical relevance. Special Issue Editorial. *Journal for Educational Research Online*, 7(2), 5–10. https://doi.org/10.25656/01:11487
- Oslund, E. L., Clemens, N. H., Simmons, D. C., & Simmons, L. E. (2018). The direct and indirect effects of word reading and vocabulary on adolescents' reading comprehension: Comparing struggling and adequate comprehenders. *Reading and Writing*, 31(2), 355–379. https://doi.org/10.1007/s11145-017-9788-3
- Paleczek, L., Seifert, S., & Gasteiger-Klicpera, B. (2017). Influences on teachers' judgment accuracy of reading abilities on second and third grade students: A multilevel analysis. *Psychology in the Schools*, *54*(3). 228–245. https://doi.org/10.1002/pits.21993
- Paleczek, L., Seifert, S., Schwab, S., & Gasteiger-Klicpera, B. (2015). Assessing reading and spelling abilities from three different angles Correlations between test scores, teachers' assessment, and children's self-assessments in L1 and L2 children. *Procedia-Social and Behavioral Sciences*, 174, 2200–2210. https://doi.org/10.1016/j.sbspro.2015.01.876
- Paloyelis, Y., Rijsdijk, F., Wood, A. C., Asherson, P., & Kuntsi, J. (2010). The genetic association between ADHD symptoms and reading difficulties: The role of inattentiveness and IQ. *Journal of Abnormal Psychology*, 38(8), 1083–1095. https://doi.org/10.1007/s10802-010-9429-7

- Pastore, S. & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003
- Penttinen, V., Pakarinen, E., & Lerkkanen, M.-K. (2020). Exploring parent-teacher trust and school involvement: A Finnish perspective. In N. K. Toren & G. J. van Schalkwyk (Eds.), *Parental involvement. Practices, improvement strategies and challenges* (pp. 65–89). Nova Science Publisher. https://doi.org/10.1080/00405840802577536
- Peterson, R. L., Boada, R., McGrath, L. M., Willcutt, E. G., Olson, R. K., & Pennington, B. F. (2017). Cognitive prediction of reading, math, and attention: Shared and unique influences. *Journal of Learning Disabilities*, 50(4), 408–421. https://doi.org/10.1177/0022219415618500
- Peura, P., Aro, T., Räikkönen, E., Viholainen, H., Koponen, T., Usher, E. L., & Aro, M. (2021). Trajectories of change in reading self-efficacy: A longitudinal analysis of self-efficacy and its sources. *Contemporary Educational Psychology*, 64, 101947. https://doi.org/10.1016/j.cedpsych.2021.101947
- Peura, P., Aro, T., Viholainen, H., Räikkönen, E., Usher, E. L., Sorvo, R., & Aro, M. (2019). Reading self-efficacy and reading fluency development among primary school children: Does specificity of self-efficacy matter? *Learning and Individual Differences*, 73, 67–78. https://doi.org/10.1016/j.lindif.2019.05.007
- Piasta, S. B., Logan, J. A. R., Farley, K. S., Strang, T. M., & Justice, L. M. (2021). Profiles and predictors of children's growth in alphabet knowledge. *Journal of Education for Students Placed at Risk*. https://doi.org/10.1080/10824669.2021.1871617
- Pierce, R., Chick, H., & Gordon, I. (2013). Teachers' perceptions of the factors influencing their engagement with statistical reports on student achievement data. *Australian Journal for Educational Research*, *57*(3), 237–255. https://doi.org/10.1177/0004944113496176
- Popham, J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. https://doi.org/10.1080/00405840802577536
- Poskiparta, E., Niemi, P., & Lepola, J. (1994). *Diagnostiset testit 1. Lukeminen ja kirjoittaminen*. [Diagnostic Tests 1, Reading and Writing]. Turun yliopisto, Oppimistutkimuksen keskus. University of Turku, Centre of Learning Research.
- Prochnow, J. E., Tunmer, W. E., & Chapman, J. W. (2013). A longitudinal investigation of the influence of literacy-related skills, self-perceptions, and inattentive behaviour on the development of literacy learning difficulties. *International Journal of Disability, Development and Education,* 60(3), 185–207. https://dx.doi.org/10.1080/1034912X.2013.812188
- Pulkkinen, J., & Jahnukainen, M. (2016). Finnish reform of the funding and provision of special education: The views of principals and municipal

- education administrators. *Educational Review, 68*(2), 171–188. https://doi.org/10.1080/00131911.2015.1060586
- Pulkkinen, J., Räikkönen, E., Jahnukainen, M., & Pirttimaa, R. (2020). How do educational reforms change the share of students in special education? Trends in special education in Finland. *European Educational Research Journal*, 19(4), 364–384. https://doi.org/10.1177/1474904119892734
- Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & di Covella, L. S. (2018). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience & Behavioral Reviews, 84,* 434–452. https://doi.org/10.1016/j.neubiorev.2017.08.001
- Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, *16*(3), 187–217. https://doi.org/10.1080/10888438.2010.538780
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, 10(1), 39–60. https://doi.org/10.17011/ht/urn.201405281859
- Roskos, K., & Neuman, S. B. (2012). Formative assessment: Simply, no additives. *The Reading Teacher*, 65(8), 534–538. https://doi.org/10.1002/TRTR.01079
- Savolainen, P. A., Timmermans, A. C., & Savolainen, H. K. (2018). Part-time special education predicts students' reading self-concept development. *Learning and Individual Differences*, *68*, 85–95. https://doi.org/10.1016/j.lindif.2018.10.005
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. [Teacher diagnosis and diagnostic competence]. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31(2), 154–165. https://doi.org/10.25656/01:13843
- Shapiro, E. S., Gebhardt, S., Flatley, K., Guard, K. B., Fu, Q., Leichman, E. S., Calhoon, M. B., & Hojnoski, R. (2017). Development and validity of the rating scales of academic skills for reading comprehension. *School Psychology Quarterly*, 32(4), 509–524. https://doi.org/10.1037/spq0000193
- Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (Specific Reading Disability). *Biological Psychiatry*, *57*(11), 1301–1309.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press. Retrieved January, 23, 2022 from https://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/detail.action?docID=3375534
- Snowling, M. J., Duff, F., Petrou, A., & Schiffeldrin, J. (2011). Identification of children at risk of dyslexia: The validity of teacher judgments using 'Phonic Phases'. *Journal of Research in Reading*, 34(2), 157–170. https://doi.org/10.1111/j.1467-9817.2011.01492.x
- So, W. W. M., & Lee, T. T. H. (2011). Influence of teachers' perceptions of teaching and learning on the implementation of Assessment for Learning inquiry study. *Assessment in Education: Principles, Policy & Practice, 18*(4), 417–432. https://doi.org/10.1080/0969594X.2011.577409

- Soares, N., Evans, T., & Patel, D. R. (2018). Specific learning disability in mathematics: A comprehensive review. *Translational Pediatrics*, 7(1), 48–62. https://doi.org/10.21037/tp.2017.08.03
- Solheim, O. J., Torppa, M., Uppstad, P. H., & Lerkkanen, M.-K. (2020). Screening for slow reading acquisition in Norway and Finland a quest for context specific predictors. *Scandinavian Journal of Educational Research*, 65(4), 584–600. https://doi.org/10.1080/00313831.2020.1739130
- Soodla, P., & Kikas, E. (2010). Teachers' judgment of students' reading difficulties and factors related to its accuracy. In A. Toomela (Ed.), *Systemic person-oriented study of child development in early primary school* (pp. 73–94). Peter Lang.
- 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
- Soodla, P., Tammik, V., & Kikas, E. (2021). Is part-time special education beneficial for children at risk for reading difficulties? An example from Estonia. *Dyslexia*, 27(1), 126–150. https://doi.org/10.1002/dys.1643
- Soodla, P., Torppa, M., Kikas, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2019). Reading comprehension from grade 1 to 6 in two shallow orthographies: Comparison of Estonian and Finnish students. *Compare*, 49(5), 681–699. https://doi.org/10.1080/03057925.2018.1445963
- Spanoudis, G. C., Papadopoulos, T. C., & Spyrou, S. (2019). Specific language impairment and reading disability: Categorical distinction or continuum? *Journal of Learning Disabilities*, 52(1), 3–14. https://doi.org/10.1177/0022219418775111
- Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal*, 111(4), 585–607. https://doi.org/10.1086/659032
- Ståhlberg, L., Hotulainen, R., & Lehto, J. (2020). Lukuseulan toimivuus ja luotettavuus peruskoululaisten lukutaidon arvioinnissa. [Validity and reliability of Read Grader in evaluating pupil's reading skills]. *NMI-Bulletin*, 30(3), 67–84.
- Stiggins, R. J. (1991). Assessment literacy. Phi Delta Kappan, 72(7), 534-539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan,* 77(3), 238–245.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan*, *83*(10), 758–765. https://doi.org/10.1177/0031721702083001010
- Sullanmaa, J., Pyhältö, K., Pietarinen, J., & Soini, T. (2019). Differences in state-and district-level stakeholders' perceptions of curriculum coherence and school impact in national curriculum reform. *Journal of Educational Administration*, *57*(3), 210–226. https://doi.org/10.1108/JEA-08-2018-0153

- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. https://doi.org/10.1037/a0027627
- Südkamp, A., Praetorius, A.-K., & Spinath, B. (2018). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*, 76, 204–213. https://doi.org/10.1016/j.tate.2017.09.016
- Takala, M., & Ahl, A. (2014). Special education in Swedish and Finnish schools: Seeing the forest or the trees? *British Journal of Special Education*, 41(1), 59–81. https://doi.org/10.1111/1467-8578.12049
- Takala, R., Pirttimaa, R., & Törmänen, M. (2009). Inclusive special education: The role of special education teachers in Finland. *British Journal of Special Education*, *3*, 162–172. https://doi.org/10.1111/j.1467-8578.2009.00432.x
- Thiede, K. W., Brendefur, J. L., Carney, M. B., Champion, J., Turner, L., Stewart, R., & Osguthorpe, R. D. (2018). Improving the accuracy of teachers' judgments of student learning. *Teaching and Teacher Education*, 76, 106–115. https://doi.org/10.1016.j.tate.2018.08.004
- Titley, J. E., D'Amato, R. C., & Koehler-Hak, K. M. (2014). Utilizing teacher ratings of student literacy to identify at-risk students: An analysis of data from the early childhood longitudinal study. *Contemporary School Psychology*, *18*, 59–68. https://doi.org/10.1007/s40688-013-0005-3
- Tolar, T. D., Fuchs, L., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability. *Journal of Learning Disabilities*, 49, 240–256. https://doi.org/10.1177/0022219414538520
- Tong, X., & Deacon, S. H. (2017). Understanding poor comprehenders in different orthographies: Universal versus language-specific skills. *Journal of Research in Reading*, 40(2), 119–124. https://doi.org/10.1111/1467-9817.12112
- 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., Niemi, P., Lerkkanen, M.-K., & Poikkeus, A.-M. (2017). The precursors of double dissociation between reading and spelling in transparent orthography. *Annals of Dyslexia*, *67*(1), 42–62. https://doi.org/10.1007/s11881-016-0131-5
- Torppa, M., Poikkeus, A.-M., Laakso, M.-L., 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., 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
- Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32. https://doi.org/10.1016/j.edurev.2020.100374
- Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal*, *56*(3), 792–821. https://doi.org/10.3102/0002831218803891
- Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2018). Teachers' high-stakes decision making. How teaching approaches influence rational and intuitive data collection. *Teaching and Teacher Education*, 71, 108–119. https://doi.org/10.1016/j.tate.2017.12.011
- Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research and Practice*, 29(2), 46–53. https://doi.org/10.1111/ldrp.12031
- Vitikka, E., Krokfors, L., & Hurmerinta, E. (2012). The Finnish national core curriculum: Structure and development. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp. 83–96). Sense Publishers. https://doi.org/10.1007/978-94-6091-811-7
- Wagner, D. L., & Espin, C. A. (2015). The reading fluency and comprehension of fifth- and sixth-grade struggling readers across brief tests of various intervention approaches. *Reading Psychology*, *36*(7), 545–578. https://doi.org/10.1080/02702711.2014.927812
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A, & Pearson, N. A. (2009). *TOSREC:* Test of sentence reading efficiency and comprehension. Pro-Ed.
- Wanzek, J., & Vaughn, S. (2009). Students demonstrating persistent low response to reading intervention: Three case studies. *Learning Disabilities Research & Practice*, 24, 151–163. https://doi.org/10.1111/j.1540-5826.2009.00289.x
- Watson, S. M. R., & Gable, R. A. (2013). Unraveling the complex nature of mathematics learning disability: Implications for research and practice. *Learning Disability Quarterly*, 36(3), 178–187. https://doi.org/10.1177/0731948712461489
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14. https://doi.org/10.1016/j.stueduc.2011.03.001
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65. https://doi.org/10.1080/0969594042000208994
- Willcutt, E. G. (2018). ADHD and reading disorder. In T. Banaschewski, D. Coghill, & A. Zuddas (Eds.), *Oxford textbook of attention deficit hyperactivity disorder* (pp. 273–279). Oxford University Press.

- Willcutt, E. G., McGrath, L. M., Pennington, B. F., Keenan, J. M., DeFries, J. C., Olson, R. K., & Wadsworth, S. J. (2019). Understanding comorbidity between specific learning disabilities. *New Directions for Child and Adolescent Development*, 165, 91–109. https://doi.org/10.1002/cad.20291
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256. https://doi.org/10.1007/s13384-013-0089-9
- Woolley, G. (2008). The assessment of reading comprehension difficulties for reading intervention. *Australian Journal of Learning Difficulties*, 13(1), 51–62. https://doi.org/10.1080/19404150802093729
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010
- Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2020). Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A five-year longitudinal study. *Child Development*, 91(1), 7–27. https://doi.org/10.1111/cdev.13123



ORIGINAL PAPERS

I

TEACHERS' ABILITY TO IDENTIFY CHILDREN AT EARLY RISK FOR READING DIFFICULTIES IN GRADE 1

by

Riitta Virinkoski, Marja-Kristiina Lerkkanen, Leena Holopainen, Kenneth Eklund, & Mikko Aro 2018

Early Childhood Education Journal, 46(5), 497-509

https://doi.org/10.1007/s10643-017-0883-5

Reproduced with the kind permission of Springer.



Teachers' Ability to Identify Children at Early Risk for Reading Difficulties in Grade 1

Riitta Virinkoski¹ · Marja-Kristiina Lerkkanen⁴ · Leena Holopainen⁵ · Kenneth Eklund² · Mikko Aro³

Published online: 5 October 2017

© Springer Science+Business Media, LLC 2017

Abstract The aim of the study was to investigate what kinds of assessment practices classroom teachers and special educational needs (SEN) teachers use in assessing first grade students' pre-reading skills (letter knowledge and phonological skills). Further, we investigated to what extent teachers were able to identify difficulties in pre-reading skills of the children with the lowest achievement scores. The accuracy of teacher ratings of students' pre-reading skills was studied by comparing teacher ratings to actual test scores. The data from two Finnish longitudinal studies were used: JLD sample (class teachers, n = 91; SEN teachers, n = 51; 200 students) and First Steps sample (class teachers, n = 136; SEN teachers, n = 34; 598 students). Results showed first, that most classroom teachers used qualitative assessment and SEN teachers also relied on tests. Secondly, although teacher ratings correlated with the test scores, closer investigation of sensitivity and specificity of the teacher ratings revealed that a number of children in need of extra support for their early reading development according to test scores remained unidentified. Moreover, there

were some students identified by the teacher to have difficulties despite test scores not confirming that. The findings underline the importance for developing more specific and reliable assessment tools for teachers to use for pedagogical purposes, and respectively, the need to pay more attention to early identification of reading difficulties in teacher training program curricula.

 $\begin{tabular}{ll} \textbf{Keywords} & Assessment \cdot Teacher \cdot Pre\text{-reading skills} \cdot \\ At\text{-risk students} \cdot Sensitivity \cdot Specificity \\ \end{tabular}$

Teachers play a key role in identifying the need for early support in reading skill development because they generally observe the first signs of reading difficulties (RD) (Bailey and Drummond 2006; Compton et al. 2010). Previous studies have indicated that teachers' judgments of reading skills in kindergarten and at the beginning of school (first and second grade) generally correspond well with the scores of standardized reading achievement test results especially regarding the high-performing students (Bailey and Drummond 2006; Begeny et al. 2011; Südkamp et al. 2012). The main purpose of teachers' evaluations of students should be to produce accurate knowledge of the students' skills in order to plan tailored instruction and support when necessary (Bailey and Drummond 2006; Mesmer and Mesmer 2008). Begeny et al. (2011) studied first- to fifth-grade students' oral reading fluency and found that accurate performance assessments could allow for providing early support, thereby preventing the need for intensive intervention. However, their findings revealed that it was difficult for teachers to judge students' reading levels as low-, average-, or highperforming. One explanation for low-judgement accuracy could be the lack of teacher training and practice in conducting assessments (Begeny et al. 2011).



Riitta Virinkoski riitta.e.virinkoski@jyu.fi

Open University, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylä, Finland

Department of Education and Department of Psychology, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylä, Finland

Department of Education, University of Jyvaskyla, P. O. Box 35, 40014 Jyvaskyla, Finland

Department of Teacher Education, University of Jyväskylä, P. O. Box 35, 40014 Jyväskylä, Finland

School of Educational Sciences and Psychology, University of Eastern Finland, P. O. Box 111, 80101 Joensuu, Finland

Particularly, children with poor pre-reading skills who are potentially at risk for reading difficulties (RD) should be identified as early as possible. Early recognition of risk for RD would be needed to avoid prolonged or more serious problems. Flynn and Rahbar (1998) also indicate that researchers have disagreed on whether teacher ratings or screening tests best identify children at risk for reading failure. In support of screening instruments for early identification, prior studies have shown that screening batteries and standardized achievement tests predict those at risk for reading failure better than teachers' evaluations based on, for example, rating scales, whereby the latter have tended to produce high false-negative rates (Fletcher and Satz 1984; Flynn and Rahbar 1998). Moreover, to lead to effective and early support for at-risk students, the screen must be relatively accurate, i.e., capable of distinguishing students who will subsequently have difficulties from those who will not (Johnson et al. 2009). Recent study by Catts et al. (2015) indicated that using screening batteries containing measures of skills such as letter naming fluency and phonological awareness enabled accurate identification of good and poor readers at the end of first grade. It has also been shown that using teacher judgment more with the universal screening procedures could increase the classification accuracy rates of at-risk and not-at-risk students (Compton et al. 2010; Martin and Shapiro 2011; Snowling et al. 2011).

In addition, studies comparing classroom teachers' and special educational needs (SEN) teachers' assessments for identifying at-risk students are lacking. For example, both class teachers and SEN teachers in Finland have a Master's Degree, and they receive different kinds of training to gain competence in identifying children at early risk for RD. SEN teacher training in Finland comprises theory and practice—not only related to individual and small-group instruction, application of various assessment tools, support in reading and writing, mathematics, and communication—but also in behavioral and socio-emotional challenges (Takala and Ahl 2014). Respectively, Finnish teacher preparation programs provide readiness to instruct a whole class within general education and adapt that instruction according to children's needs.

Consequently, the present study investigates classroom teachers' and SEN teachers' assessment practices and a matching of their ratings of pre-reading skills regarding especially children who have difficulties in letter knowledge and phonological skills. Further, we explored in how well the teachers' ratings correspond to the test scores at the beginning of the first grade in the highly transparent Finnish orthography.



Assessment of Pre-Reading Skills by Teachers

Previous research indicates that the most common rationale for being identified as an at-risk student is problems in, for example, letter knowledge and identifying letter sounds. To ensure accurate identification, the screening batteries should cover several skill areas related to developing reading skills, such as phonological skills, orthographic and letter knowledge, word reading ability, vocabulary, and syntactic ability (Bailey and Drummond 2006; Davis et al. 2007). To accurately classify students into two groups, at risk and not at risk for poor reading outcomes, it is important that the screens are targeted at reading skills, and that the content is age-appropriate (Jenkins et al. 2007). However, the accuracy of screening measures differs with respect to sensitivity and specificity (Catts et al. 2015; Compton et al. 2010; Jenkins et al. 2007; Johnson et al. 2009). Sensitivity refers to the degree of true positives, meaning how accurately the measure identifies students at high risk for RD. Specificity, on the other hand, refers to the degree of true negatives, or how accurately the measure identifies students at low risk for RD. The fact that a test discriminates poor readers at the group level does not necessarily guarantee accuraacy in predicting or identifying difficulties at the individual level (Puolakanaho et al. 2007). The quality of the predictor is determined by how well it is able to capture the true-positive cases that turn out to have RD at school age, and to avoid false-positive cases that predict risk for RD although the children do not have difficulties in reading at school age. According to the literature, teachers' assessment practices can be divided into three categories: tests comprising screening or individual test batteries, (performance-based assessment), curriculumbased measures (CBM), and qualitative assessments such as observations in the classroom (Bailey and Drummond 2006; Südkamp et al. 2012). One way to assess student progress toward long-term curriculum goals in literacy learning is CBM, which is the main tool of screening difficulties learning difficulties and the risk for RD in the response to intervention (RTI) framework (Stecker et al. 2005; Deno 1985). CBM may be used to monitor students' progress in an entire school or classroom, to track an individual's progress toward end-of-year benchmarks or individualized education program goals, or to screen students at a specific time point to determine their level of risk for academic failure (Deno 1985, 2003; Madelaine and Wheldall 2005; Zumeta et al. 2012).

A number of previous studies (e.g., Bailey and Drummond 2006; Beswick et al. 2005) have shown that teachers' evaluations and their perceptions of a student's risk for literacy failure can be used as early as the beginning of kindergarten and the first grade to identify signs of RD. In Bailey and Drummond's (2006) study, kindergarten and first-grade teachers were asked to identify one to four students in their

class who they perceived to be at risk for RD, but who were not receiving any formal remediation at the moment. They used literacy development checklists (LDC; Bailey et al. 2001) and also concept maps based on targeted early literacy skills, such as decoding, letter-sound correspondence and phonemic awareness. However, according to Bailey and Drummond (2006), the data teachers rely on when rating students' reading performance may not allow for making accurate judgments of particular pre-reading skills. Teachers' decisions seemed to be sometimes based on situational or other irrelevant factors (e.g., gender, behavior, students' ability to work in groups), instead of solely performance assessments (Beswick et al. 2005). They might also have insufficient knowledge or competence to identify students' RD (Bailey and Drummond 2006). In addition, Bailey and Drummond (2006) noted that some teacher characteristics, such as years of teaching experience and personality, affect the accuracy of teacher judgments. Furthermore, teachers have been shown to have a tendency to underestimate the reading skills of those students who have had prior weaknesses in reading, and whose general cognitive skills are at a low level in combination with previously identified SEN (Soodla and Kikas 2010).

Correspondence Between Teacher Ratings and Test Scores

In most studies, the correlations between teacher ratings and test scores have varied between 0.40 and 0.70. For example, Südkamp et al. (2012) meta-analysis on teachers' judgment accuracy in a regular school system, from kindergarten through grade 12 over a 20-year period, indicated that the correlation between teacher judgments of students' academic achievement in language arts (reading, spelling, literature, and composition) and mathematics, and their actual test performance was moderate, at 0.63. Their findings are in line with Hoge and Coladarci's (1989) study that investigated language arts, mathematics, and the social sciences, where the median correlation of 0.66 was reported. In their study, correlations between teacher judgments and the standardized tests ranged from 0.41 to 0.92.

However, in Bailey and Drummond's (2006) study, the correlations between teacher ratings and standardized tests regarding the emergent literacy/basic reading skills domain (e.g., print and graphic presentations) and the phonological awareness subskill of kindergarten and first-grade students were weak and not significant. They came to the conclusion that the low correlation resulted from the array of informal assessment procedures that teachers used, such as combining old curricular material with current material and observational information with in-class tests.

Despite relatively high correlations between teachers' evaluations and children's actual test scores, teachers may, however, systematically over- or underestimate student performance (Bates and Nettlebeck 2001). Another salient limitation of teacher judgments may be revealed when the range of student competence is restricted, particularly regarding students who show low-academic performance (Graney 2008; Soodla and Kikas 2010). Teachers' judgments may also be related to some personal characteristics, such as their skills, training, future expectations, or perceptions (Kikas et al. 2015; Soodla and Kikas 2010).

Flynn and Rahbar (1998) developed a theory-based screening instrument for teachers to assess reading competency, and their results suggest that teachers' predictions of children at risk for RD can be improved by using rating instruments that include research-validated antecedents of reading with behavioral descriptions of low and high achievement (Flynn and Rahbar 1998). Further, in their study, Bailey and Drummond (2006) found that, by using a literacy checklist, teacher evaluations can become more systematized and also lead to a higher identification rate of at-risk students.

The best predictors of a preschooler's or kindergartener's later reading achievement when the child has a familial history of dyslexia have turned out to be measures that require processing printed material, together with oral language proficiency measures and performance-IQ measures (Puolakanaho et al. 2007). Most studies evaluating the accuracy of teachers' judgments have used standardized tests as the comparative criterion for this investigation. Fletcher and Satz (1984) and Flynn and Rahbar (1998) found in their studies that compared to teacher ratings, standardized tests more accurately identified students who were potentially at risk for RD in the future. Teacher ratings usually had high falsenegative rates and low true-positive rates. Flynn and Rahbar (1998) also found that combining class teacher ratings and screening tests in the first, second, or third grades increased the accuracy of identifying students who would experience reading failure in the future, with a correct identification percentage of 88%. In the same study, kindergarten teachers only used a traditional rating scale to predict future reading achievement, and the positive identification rate was rather low (30%). However, in this same study, using a projectdeveloped, theory-based screening battery, the class teachers correctly identified 81% of poor readers. The prediction rate of the teacher ratings improved after some researchvalidated changes had been made, but remained below the identification accuracy of the screening test.

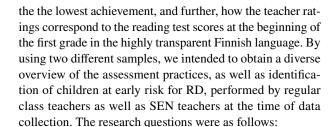


Learning to Read in Finnish

Finnish children attend kindergarten at age 6, and reading instruction begins at age 7 when they enter first grade. Upon entering school, letter knowledge seems to be one of the best predictors of reading and spelling accuracy in the Finnish language (Holopainen et al. 2001; Lerkkanen et al. 2004). Also, phoneme identification and pseudoword repetition at school entry predict the development of accuracy in reading and spelling (Aro 2006). The Finnish orthography is almost purely phonemic: the grapheme-phoneme correspondences are regular and symmetrical at the level of the single letter, and early reading instruction in Finnish is almost uniformly rests upon synthetic phonics (Aro 2006; Seymour et al. 2003). In transparent orthographies, such as Finnish, the process of learning to decode accurately is rather fast (Seymour et al. 2003), and that might make the early identification of risk for RD, manifested mostly as problems in reading rate, even more challenging for teachers. Studies have shown that approximately 30% of Finnish children are able to decode before entering the first grade (Soodla et al. 2015), and highly accurate decoding skills are usually acquired within the first months of reading instruction (Lerkkanen et al. 2004). Even the nonreaders at school entry reach the level of early readers in reading accuracy during the first school year (Lerkkanen et al. 2004; Parrila et al. 2005; Soodla et al. 2015). However, students whose growth is slow for letter knowledge and phonological awareness could encounter RD at the beginning of school (Lyytinen et al. 2006; Torppa et al. 2006). Additionally, a study identified a group of children with problems in phonological decoding at the end of the second grade, who continued to remain behind their peers in reading accuracy at grade 8 (Eklund et al. 2015). In general, studies have shown that Finnish students who struggle with reading do not typically have problems with reading accuracy, but do experience persistent problems with reading fluency (Hintikka et al. 2008). In the case of RD, the forms of support are remedial teaching during or after school by the class teacher, part-time special education given by the SEN teacher individually or in small groups during school days, or co-teaching by the class teacher and the SEN teacher during normal literacy lessons (Lerkkanen 2007). However, these forms of support do not require any formal diagnosis of a reading difficulty (Björn et al. 2016).

The Aim of the Present Study

The aim of the study was to investigate teachers' evaluation practices, and the sensitivity and specificity of their assessments of pre-reading skills, especially of children with



- (1) Which assessment practices do class teachers and SEN teachers use to assess pre-reading skills (e.g., letter knowledge, phonological skills) at the beginning of grade 1? According to previous studies (e.g., Bailey and Drummond 2006; Beswick et al. 2005; Compton et al. 2010), we expected that teachers use screening batteries, CBM, and observation when assessing pre-reading skills (Hypothesis 1). Also expected were variations between the practices used by class teachers and SEN teachers.
- (2a) Are teacher ratings associated with test scores in prereading skills? We expected the teacher ratings to correspond quite well with test scores (Hypothesis 2a, see e.g., Graney 2008; Hoge and Coladarci 1989).
- (2b) How accurately do the teachers identify students' prereading difficulties to test scores, and what are the sensitivity and specificity rates of their assessments? According to previous studies (Fletcher and Satz 1984; Flynn and Rahbar 1998), we expected that teacher ratings would have had high-false negative rates, and on the other hand, low true-positive rates in identifying at-risk students for reading. (Hypothesis 2b).

Method

The data for this study were drawn from two Finnish longitudinal studies. In both studies, parents and teachers were asked for written consent for the child's and their own participation in the study. In the Jyväskylä Longitudinal Study of Dyslexia (JLD), only the responses of SEN teachers were available concerning RQ1, but in the First Steps sample, we had the opportunity to study both regular classroom teachers' and SEN teachers' responses. Regarding RQ2, in the JLD sample, both class teachers' and SEN teachers' assessments were gathered, whereas in the First Steps sample, only SEN teachers' responses were available.



JLD Sample

Participants and Procedure

In this study, we used the data from the fall of the first grade, and the data comprised class teachers (n=91), SEN teachers (n=51), and 200 first-grade students (M age = 7.19 years, SD = 0.26; 47% girls, 53% boys). The student data comprises four successive age cohorts born between 1993 and 1996, and half of the students had a familial risk of dyslexia, and the other half belonged to the control group. The at-risk children were defined by parents' self-reports of literacy difficulties and their reports of similar problems among their immediate relatives. The parents were sent a questionnaire that dealt with demographic information, and the occurrence of reading and writing difficulties during childhood, adulthood, and among relatives. In diagnostic tasks of reading and writing, the parents selected in the dyslexic sample had to obtain a -1 or less z-score in either accuracy or speed of oral text reading or spelling accuracy. Also, they had to obtain z-scores of -1.0 or less in two or more out of eight computer-aided measures. The children belonging to the control group or low-risk group did not have any reported parental literacy difficulties nor in their first- or seconddegree relatives (for more specific details, see Leinonen et al. 2001).

The research data consisted of teachers' questionnaires, teachers' student ratings, and test scores regarding first-grade students' letter knowledge and phonological skills upon entering school. The majority of the teachers' observation forms were returned in December and some after a reminder in February. Students participated in the individual tests in August.

Measures

Special Education Teachers' Assessment Practices

SEN teachers reported their reading assessment practices by responding to an open-ended question in the SEN teacher's observation form: "What kinds of practices have you used in your assessment of learning to read and write and what subskills have you assessed?" Nineteen (37%) out of 51 SEN teachers responded to this question. The teachers' written responses of assessment practices were classified in the following categories: (1) qualitative assessment (e.g., observation, checklists, or discussions); (2) CBM (e.g., tests of ABC books or spelling from dictation); (3) reading tests, such as screening or individual tests; and (4) "Other," comprising teachers' self-developed assessment tools. Class teachers were not asked to report their assessment practices.

Teachers' Ratings Concerning Children's Reading and Pre-reading Skills

Regular classroom teachers and SEN teachers rated the students' reading performance on a five-point scale: 1 = "clear problem"; 2 = "mild problem"; 3 = "masters the skill"; 4 = "masters the skill quite well"; and 5 = "masters the skill very well." For this study, we selected three pre-reading skills rated by both class teachers and SEN teachers: letter naming, initial phoneme identification, and blending three sounds. The categories "clear problem" and "mild problem" were pooled together for the analyses; in addition, the categories "masters the skill," "masters the skill quite well," and "masters the skill very well" represented in the analyses that the student did not have any problem with the skill from the teachers' point of view.

Letter Knowledge

The letter-naming task was administered as an individual test in August. Twenty-nine uppercase letters were shown to the student by a trained tester on a sheet of paper in a fixed order, and the student was asked to name them as accurately and quickly as possible. The score was based on the number of correct responses.

Phonological Skills

The phonological skills were assessed with two individual tasks drawn from Diagnostic Test Battery 1 (Poskiparta et al. 1994). In the initial phoneme identification task, the trained tester first said the word to the student, and then the student said which sound was at the beginning of the word. In the phoneme-blending task, the experimenter said altogether 10 word items, phoneme by phoneme, and the student was instructed to say the resulting word. The sum scores were based on the number of correct items. Cronbach's alpha of phoneme identification was 0.94. and phoneme blending was 0.57. To enhance the reliability of the phonological awareness task, these variables were merged, and the mean of the two tasks was 0.80.

First Steps Sample

Participants and Procedure

In this study, the data comprised regular classroom teachers (N=136; M age = 42.69 years, SD=9.1; 91% female, 9% male), SEN teachers (N=34; M age = 45.62, SD=9.60; 97% female, 3% male), and a subsample of 598 children selected for more intensive follow-up (47% girls, 53% boys) from four municipalities participating in the study in the fall of



the first grade. The children were 7 years old at the beginning of the first grade (beginning of school; M=7.18 years, SD = 0.30). The large majority of the class teachers (78%) had a master's degree in education from a class teacher program (69%). The rest had a master's degree in either special education (5%) or both programs (4%). One percent of the class teachers had a doctoral degree in education. Twentyone percent (21%) had some other degree, typically a bachelor's (BA) degree, which was formerly sufficient for the qualification as a class teacher or kindergarten teacher. The basic education for SEN teachers was a master's degree from a teacher education program combined with an SEN teacher qualification (53%), a master's degree from an SEN teacher program (44%), or something else (3%), such as a BA degree as a kindergarten teacher. Two class teachers and two SEN teachers did not report their education.

The sample of the present study contained both students identified as at risk (n = 277) and not at risk (n = 321) for RD. From the total sample of 1,880 children, the children's risk for reading problems was determined by the researchers at the end of the kindergarten year on the basis of four criteria: children's initial phoneme identification (indicator of phonological awareness), letter knowledge, rapid automatized naming, and parental report of their own RD (see Lerkkanen et al. 2011). The risk for RD was defined as a joint occurrence of at least two criteria out of three: low phonological awareness (i.e., scored clearly below age level in initial phoneme identification, $\leq 15\%$); poor letter knowledge (\leq 15%); and poor rapid automatized naming (\leq 15%; Kiuru et al. 2013; Lerkkanen et al. 2011). Furthermore, if parents reported having reading disabilities, a score below the 15th percentile in one of the three tests (phonological awareness, letter knowledge, or rapid automatized naming) was sufficient for identifying a risk for RD. The control children were randomly selected from the same classrooms as the children identified as being at risk for RD. The criteria resulted in one to six (typically two or three) children from each participating classroom being included in the more intensive follow-up. One to five from a maximum of six children were from the at-risk group (depending on the number of at-risk children in the classroom in each case), and the remainder were from the no-risk group.

SEN teachers were sent a list of the students who were followed more intensively in their school, but they did not know which group (at risk for RD or not) the individual children belonged to. They were asked to rate all the students who had received part-time special education during the first grade in that particular school, irrespective of the reason for support (e.g., speech therapy, reading, math, behavioral problems) by December. In some cases, if the number of students exceeded six, the SEN teacher was allowed to select the students for the rating (usually students who needed more intensive support were selected).

The individual and the group tests at the beginning of the first grade were carried out in September. If the student was absent for the tests in September, the tests were implemented in October.

Measures

Regular Classroom Teachers' and SEN Teachers' Assessment Practices

In December, both class teachers and SEN teachers reported the assessment practices they used with their students by answering the question on the SEN teachers' questionnaire "How has the need for special educational support been defined? (What was assessed?/How was it assessed?/When did the assessment take place?)?" The teachers' responses were classified in three categories similarly to the JLD sample: qualitative assessments, CBM, and tests.

Ratings by SEN Teachers

SEN teachers were asked to evaluate their students' school entry pre-reading (e.g., letter knowledge) skills in December by filling in questionnaires concerning individual children. They rated the students' pre-reading skills using a three-point rating scale: 1="clear problem"; 2="mild problem"; and 3="no problem." Two variables were selected from the questionnaire for this study: letter naming and phonological skills (reading/spelling 3–4-letter syllables). The categories "clear problem" and "mild problem" were pooled together for the analysis because we were only interested in whether or not the student had difficulty from the teacher's point of view.

Letter Knowledge

Letter knowledge was assessed in an individual situation using the ARMI test battery (Lerkkanen et al. 2006). The children were instructed to name 29 letters of the Finnish alphabet arranged randomly in three rows. The score was the number of correctly named letters (max = 29). Cronbach's alpha for the naming letters task was 0.92.

Phonological Skills

The phoneme-blending task (Poskiparta et al. 1994) was a group-administered test. The experimenter said words phoneme by phoneme, and after each word, the students were shown four pictures of objects, from which they had to choose the picture similar to the word formed by the phonemes. The score was the sum of correct items (maximum score 9). Cronbach's alpha was 0.70.



Table 1 Number of assessment practices of the SEN teachers in the JLD sample (n=19)

Number of assessment practices	n	%
One assessment practice	10	53
Two assessment practices	4	21
Three assessment practices	5	26
Total	19	100

Table 2 The assessment practices of class teachers and the SEN teachers in the first steps sample

Number of assessment practices	Class teachers		SEN teachers	
	n	%	n	%
One assessment practice	17	70	10	37
Two assessment practices	7	30	15	56
Three assessment practices	0	0	2	7
Total	24	100	27	100

Data Analyses

The first research question was examined using descriptive statistics, and the analyses of the second research question were carried out using Spearman's rank-order correlation and cross tabulations. Next, we calculated the sensitivity and specificity scores of the teacher ratings in order to assess the overall accuracy of the teacher ratings with regard to identifying an early risk for problems. The cut-off score for low achievement in the test data was set to the lowest 15th percentile. Finally, logistic regression analyses were conducted to test statistically whether the teachers' ratings of students' pre-reading skills were significantly interconnected with the dichotomized students' test scores.

Results

Teachers' Assessment Practices

First, we examined the assessment practices teachers used to evaluate pre-reading skills (letter knowledge and phonological skills) at the beginning of grade 1.

JLD Sample

The number of assessment practices used by SEN teachers are summarized in Table 1. Most SEN teachers reported that they used only one type of assessment, qualitative or CBM being the most commonly used. If the SEN teachers used two types of assessment practices, they were usually tests combined with qualitative assessment. Further, when the SEN teachers used three assessment practices, the most common combination was tests, qualitative assessment, and CBM. Altogether, tests were used by 47% of the SEN teachers.

First Steps Sample

Class teachers' and SEN teachers' assessment practices in the First Steps sample are summarized in Table 2. Assessment practices were unevenly distributed in the two groups $(\chi^2 (2, N=51)=6.57, p<.05)$. According to the standardized residuals class teachers used more often and SEN teachers less often than expected only one assessment practice (adjusted standardized residual for the cells = 2.4). Moreover, the use of two assessment practices was close to significant in favor of SEN teachers the adjusted standardized residual being 1.9. Most class teachers (58%) used qualitative assessment as their only practice. Further, when the class teachers assessed students using two types of practices, they were usually either CBM or tests combined with qualitative assessment. More than half of the SEN teachers relied on two types of assessment practices, most commonly tests

 Table 3 Correlations between class teacher and SEN teacher ratings and test scores in the JLD sample

Class teachers' ratings	Test scores (n=40-44)		SEN teachers' ratings	Test scores $(n=34-35)$	
	Letter knowledge	Phonological awareness ^a		Letter knowledge	Phono- logical awareness ^a
Letter knowledge	0.42**	0.41**	Letter knowledge	0.52**	0.35*
Phoneme identification	0.43**	0.33*	Phoneme identification	0.45**	0.46**
Phoneme blending	0.52**	0.30	Phoneme blending	0.49**	0.32

^{*}Correlation is significant at the 0.05 level (two-tailed)

^aThe phonological awareness test variable comprises the variables of initial phoneme identification and blending phonemes



^{**}Correlation is significant at the 0.01 level (two-tailed)

Table 4 Correlations between the SEN teacher ratings and the test scores in the first steps sample

	Test scores			
SEN teachers' ratings	Letter knowledge (n = 69)	Phoneme blending (n = 69)		
Letter knowledge	0.50**	0.17		
Phoneme blending ^a	0.24	0.29*		

^{*}Correlation is significant at the 0.05 level (two-tailed)

combined with CBM. None of the class teachers and only 7% of the SEN teachers reported using all three types of assessment. Nearly 90% of the SEN teachers used tests in their assessment, either tests only, or tests with some other assessment practice. Whereas among the class teachers, only 13% reported that they used tests as their only assessment practice or combined them with qualitative assessment.

Association Between the Teacher Ratings and the Test Scores

Regarding our second research question, we wanted to determine the associations between teacher ratings and reading test scores, especially of the lowest achieving students. Spearman's correlations (see Table 3 for the JLD sample and Table 4 for the First Steps sample) showed that associations between teachers' ratings and the test scores were moderate. The letter-knowledge task in the JLD sample correlated quite well (0.52) between the SEN teachers' letter-knowledge ratings and the class teachers' phoneme-blending ratings. The

best correlation in the First Steps sample appeared in the letter-knowledge task (0.50).

Finally, we analyzed how accurately the class teachers and the SEN teachers were able to identify students at risk for reading failure (sensitivity), and on the other hand, those who were not at risk (specificity).

JLD Sample

In letter naming, the cut-off score used to indicate problems was 19 correct letters out of 29 in the individual letter-naming task, and 31 students scored below this score. In the phoneme-identification task, the cut-off score was 3 or fewer correct answers out of 10 phonemes, and there were 30 students in this group. Further, in the phonemeblending task, the lowest achieving students had 1 or 0 correct responses out of 10 items in the test (the cut-off score), and there were 34 students in this group. Tables 5 (class teachers) and 6 (SEN teachers) present the true positives, the false negatives, the true negatives, and the false positives, according to the test scores. According to logistic regression analysis the class teachers' ratings of students' letter knowledge and the students' categorical test scores were close to significant (χ^2 (1) = 2.80, p = .09). In addition, class teachers' ratings of students' phonological awareness were not associated with their categorical test scores (χ^2 (1) = 0.90, p = .34 and χ^2 (1) = 1.20, p = .27 for phoneme identification and phoneme blending, respectively).

The sensitivity of class teacher ratings in letter knowledge was 31% and specificity was 90%, which means that 69% of the at-risk students remained unidentified, and 10% of the students with no difficulties were falsely identified as at-risk. In phoneme identification, the sensitivity rate was 17% and the specificity rate was 93%, which reflects the fact that, in

Table 5 Identification of students at risk for RD based on the class teacher ratings and the test scores in the JLD sample

Pre-reading skill (n=number of students rated by class teacher)	True positives n (%)	False negatives n (%)	True negatives n (%)	False positives n (%)
Letter knowledge (n=44)	4 (9)	9 (21)	28 (63)	3 (7)
Phoneme identification $(n=42)$	2 (5)	10 (24)	28 (66)	2 (5)
Phoneme blending ^a $(n=42)$	7 (17)	8 (19)	19 (45)	8 (19)

^aThe phoneme-blending task corresponds to blending three sounds in the class teachers' ratings

Table 6 Identification of students at risk for RD based on the SEN teacher ratings and the test scores in the JLD sample

Pre-reading skill (n=number of students rated by teacher)	True positives n (%)	False negatives n (%)	True negatives n (%)	False positives n (%)
Letter knowledge (n=36)	10 (28)	8 (22)	15 (42)	3 (8)
Phoneme identification $(n=33)$	6 (18)	8 (24)	16 (49)	3 (9)
Phoneme blending ^a $(n=35)$	8 (23)	3 (9)	13 (37)	11 (31)

^aThe phoneme-blending test variable corresponds to blending three sounds in the SEN teachers' ratings



^{**}Correlation is significant at the 0.01 level (two-tailed)

^aThe phoneme-blending test corresponds to reading/writing 3–4-letter syllables

general, teachers very rarely identified problems in phoneme identification. Finally, in phoneme blending, the sensitivity rate was 46% and the specificity rate was 70%, which indicates that the class teachers did not identify about half of the at-risk students; additionally, they identified 30% of the notat-risk students as having difficulties in phoneme blending. The results indicate that it was very challenging for the class teachers to identify the difficulties, in general; albeit in phoneme blending, the ratings were more in line with the test scores. According to logistic regression analysis the SEN teachers' ratings of students' letter knowledge were associated with students' categorical letter knowledge test scores $(\chi^2(1) = 5.6$. p = .018). Regarding phoneme identification the SEN teachers' ratings and students' categorical test scores were close to significant (χ^2 (1)=3.0, p=.08), and in phoneme blending the SEN teachers' ratings and the categorical test scores were not associated (χ^2 (1) = 2.3, p = .13).

The sensitivity of SEN teacher ratings in letter knowledge was 55% and specificity was 83%, which means that about half of the at-risk students were identified, but also 17% of the not-at-risk students, according to the tests, were unnecessarily identified. In phoneme identification the sensitivity rate was 43% and specificity rate 84%. This shows that SEN teachers had difficulties especially in recognizing the at-risk students struggling with phoneme identification. As in phoneme blending, the sensitivity rate was 72% and specificity was 54%. These results indicate that the majority of the atrisk students were identified, but also the rate of unnecessarily recognized students was quite high. These results show that it was also challenging for the SEN teachers to identify at-risk students who had difficulties with phonological skills. However, the SEN teachers seemed to identify RD more than the class teachers, and somewhat more accurately. Nonetheless, they also missed most students who were having difficulties.

First Steps Sample

To score below the cut-off point for low achievement in letter knowledge, the student had to correctly name a maximum of 14 out of 29 letters, and there were 85 students in this group. The SEN teachers had rated 26 of those students in the letter-knowledge task. If in the phoneme-blending task, the student got a maximum of 5 correct answers out of 10, the student

belonged to the lowest-achieving group. The number of students who scored below this cut-off score was 114, and the SEN teachers had rated 24 of those students.

Table 7 presents the true positives, the false negatives, the true negatives, and the false positives in the First Steps sample.

Regarding the First Steps sample, the results first showed that sensitivity of the SEN teacher ratings for letter knowledge was 100% and specificity was 23%, which means that all at-risk students were identified; however, 77% of the students were identified as at-risk even though, according to their test scores, they did not have difficulties with letter knowledge. Further, the sensitivity of teacher ratings for phoneme blending was 100%, whereas specificity was only 9%. Thus, the SEN teachers identified all at-risk students, but they also estimated that 91% of the students who managed quite well in the tests had difficulties with phoneme blending. According to logistic regression analyses and letter knowledge the SEN teachers' ratings were highly associated with the students' categorical letter knowledge test scores (χ^2 (1) = 10.46, p = .001), and also in phoneme blending to some extent $(\chi^2(1) = 3.5, p = .06)$.

Discussion

The aim of this study was to get answers to three research questions. First, we wanted to describe the assessment practices the teachers used in identifying difficulties in students' pre-reading skills (letter knowledge and phonological skills) upon entering school in the first grade. The results first showed that the class teachers mostly used one single assessment practice, whereas the SEN teachers often used a combination of several assessment practices. Second, it turned out that the correlations between teacher ratings and test scores were mostly weak or moderate. In addition, we studied the accuracy of the class teachers' and SEN teachers' ability to identify the lowest achievers based on the test scores. To investigate this, we counted the sensitivity and specificity of the ratings. For the JLD sample, there were differences between the accuracy of the class teachers' and the SEN teachers' ratings, and in the First Steps sample, the specificity rate, in particular, was very low.

Table 7 Identification of students at risk for RD based on the SEN teachers' ratings and the test scores in the first steps sample

Pre-reading skill (n=69, number of students rated by SEN teachers)	True positives n (%)	False negatives n (%)	True negatives n (%)	False positives n (%)
Letter knowledge	26 (38)	0 (0)	10 (14)	33 (48)
Phoneme blending ^a	24 (35)	0 (0)	4 (5)	41 (60)

^aThe phoneme-blending test corresponds to reading/spelling 3-4-letter syllables in the SEN teachers' ratings



First, we were interested in finding out the kinds of assessment tools the teachers used to evaluate students' pre-reading skills. We expected (Hypothesis 1) that all teachers would have used versatile assessment practices (Graney 2008). Instead, most class teachers relied on qualitative assessment, unlike the SEN teachers. A minority of the SEN teachers reported that they used qualitative assessment solely, and a few combined qualitative assessment with some other means of assessment. It has been shown (Bailey and Drummond 2006; Martin and Shapiro 2011) that the qualitative data sometimes used by teachers is not sufficiently accurate or reliable for making decisions on particular skills.

Further, contradictory to what was expected (Hypothesis 2a), the correlations between the teacher ratings and the actual test scores were significant but mostly moderate. The main reason for this finding might be that the teachers had rated the students' skills with 3- and 5-point scales, and the test scores were continuous variables. In previous studies (Feinberg and Shapiro 2003; Flynn and Rahbar 1998), the rating scales and instruments have been more consistent with each other. Our study is also in line with Südkamp et al. (2012), who found that achievement tests usually measure very specific areas of academic ability, while teachers' ratings can be much broader evaluations of a skill (e.g., overall ability in reading). Additionally, according to previous research (Flynn and Rahbar 1998; Martin and Shapiro 2011; Speece et al. 2011), teacher ratings combined with screening tests have proven to be the most accurate instrument for detecting students who might later confront RD. For example, in Flynn and Rahbar's (1998) study, 88% of at-risk students were discovered by combining both methods.

Finally, partly as we expected (Hypothesis 2b), there were high false-negative rates in both class teachers' and SEN teachers' ratings (JLD sample). Also, the true-positive rate was low in the class teachers' ratings in the JLD sample (Fletcher and Satz 1984; Flynn and Rahbar 1998). Contradictory to what was expected (Hypothesis 2b), in the First Steps sample, the true-positive rate was high, but remarkably, the false-positive rate was also extremely high. One explanation might be that teachers are more used to evaluating more comprehensively students' reading and writing skills, instead of specific sub-skills. It could also be difficult for SEN teachers to recognize when the student no longer needs support or how well the student's skills have developed. Perhaps this finding can be explained by the fact that the SEN teachers in this study only rated those students who had previously received support for their learning, and not necessarily RD (see Soodla and Kikas 2010).

The current study differs from previous studies in that both class teachers' and SEN teachers' data were available. This enabled, to some extent, drawing comparisons between the two teacher groups. According to this study (JLD sample), the SEN teachers appeared to identify atrisk students a bit more accurately than the class teachers, because their valid positive rate was higher. An explanation for this might be that SEN teachers' have more opportunities to evaluate students and are also in a better position to support individual students than class teachers. Also, SEN teacher education may provide SEN teachers with greater competency and knowledge to use various assessment tools in their work, compared to class teachers.

A key finding in this study was that there was only a weak link between the teachers' ratings and the test scores. Both underestimations and overestimations of the difficulties were made, especially by the class teachers (JLD sample). Also, in the First Steps sample, the SEN teachers identified significantly more difficulties in pre-reading skills than the students actually had, according to their test scores. The SEN teachers' assessments could have been influenced by the fact that those students had previously received part-time special education for some learning difficulties (Soodla and Kikas 2010).

Some questions still need to be discussed. First, are the teachers' assessment practices sensitive enough so that most, if not all, of the students in need for support can be detected by using them? In addition, could using several assessment practices improve the accuracy of teacher ratings? According to our findings, at least some SEN teachers have assessed the students for difficulties in pre-reading skills using several assessment practices. Unfortunately in this study we could not show whether use of multiple practices had resulted to more accurate identification of RD. Anyway, using multiple assessment practices could enable teachers to provide targeted and individually designed support measures to improve a certain skill when a difficulty is carefully defined. However, this study also shows that, at times, the SEN teachers had evaluated the students' skills using multiple practices, even though the students' test scores were above the cut-off scores. Thus, there is discrepancy between the SEN teachers' perceptions and the actual test scores. Second, this study raises the question of reliability and stability of the test results, as well as the teachers' ratings, especially regarding those students who had been identified as false positives at the beginning of the first grade. One longitudinal study has indicated that late-emerging dyslexia seems rather difficult to predict (Torppa et al. 2015). In this study (the First Steps sample), most SEN teachers used tests to assess their students' skills, either alone or with some other assessment practice. That might be the starting point for further and more specific investigation of the difficulty, using additional assessment practices. The need to better understand teachers' impressions stems from research showing that information from formal screening tests and teacher ratings together increase the accuracy of detecting RD in the early



elementary grades (Bailey and Drummond 2006; Flynn and Rahbar 1998; Martin and Shapiro 2011).

Finally, does the high rate of false positives lead to the fact that teachers are giving support to students who may be able to learn to read quite well without support, and instead, some of the at-risk students are not getting the support they need? Fletcher and Satz (1984) suggested in their study that students identified as at-risk could be included in classroombased small-group interventions targeted to the skill deficits identified by the screening battery. Working with these small groups, teachers could reassign children who progress rapidly to other activities, while continuing to intervene with those who struggle with their reading. This kind of flexible teaching and support model is already being used in Finland (see Lerkkanen 2007; Björn et al. 2016), when the class teacher and the SEN teacher work together in the classroom. According to this study, most class teachers used only qualitative assessment, which is a parallel finding with previous studies (e.g., Bailey and Drummond 2006). For this reason, we see that collaboration between class teachers and SEN teachers on assessment issues is desirable, if not necessary.

Teachers have a unique position for the early identification of students' RD, and this requires expertise, as well as the appropriate assessment tools. In order to be able to identify at-risk students and to deliver effective support and interventions in reading, SEN teachers, as well as class teachers, must be able to recognize students' deficits accurately and as early as possible. Early identification, and also intervention in specific deficit areas, can improve students' reading skill levels immediately, as well as prevent later difficulties. The results of this study indicate that teachers need reliable tools, not only to identify difficulties, but also to follow-up on skill development.

Limitations

Before drawing any generalizations from the findings, there are some limitations that should be highlighted. First, in both samples there were missing data, and accordingly, the comparison between class teachers' and SEN teachers' data, for example, was rather complicated. Further, the rather small sample of regular class teachers and SEN teachers did not allow for an analysis of teachers' assessment practices and their relationship to the accuracy of their judgements. In both samples, the teachers were also aware of the fact that there were more students with difficulties among the samples than there would have been if the sample had been based on unselected samples. Thus, it is possible that, in some cases, the teachers assumed the student had difficulties in reading, based on their prior knowledge about the student's low achievement. Furthermore, the variables in the tests and the teachers' ratings (i.e., what teachers were asked to assess) were not entirely comparable to each other.

Conclusions

The results of this study add to our understanding of class teachers' and special education teachers' essential role, and also their ability to evaluate students' pre-reading skills at the beginning of the first grade. The present study revealed that SEN teachers were able to quite accurately identify students at-risk for RD, however they seemed to face challenges in monitoring the progress in their students' literacy skills. Apart from identifying the need for support at early stage of learning to read it is also as important to evaluate students' development of literacy skills using dynamic assessment practices. That could help the SEN teachers to decide when some student no longer is in need for support, and they could have more resources in supporting the at-risk students. Our findings suggest that more attention should be paid to teacher training, as well as developing reliable assessment tools for teachers. Especially, every teacher's expertise in various assessment practices for the early identification of students at risk for RD should be ensured. Further, the current findings emphasize the need for developing high-quality tools that would also enable a systematic and reliable followup of a student's skills.

Acknowledgements This study has been carried out in the Centre of Excellence in Learning and Motivation Research, and financed by the Academy of Finland (No. 213486 for 2006–2011) and other grants from the same funding agency for the authors (Nos. 292466 for 2015–2019, 268586 for 2013–2017).

References

Aro, M. (2006). Learning to read: The effect of orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 531–550). Mahwah, NJ: Lawrence Erlbaum Associates.

Bailey, A. L., Cano, L., Fischer, D., Freeman, S., Jacobs, J., Heritage, M., et al. (2001). The LDC manual: A guide to using the Literacy Development Checklist (Rev. ed.). Los Angeles: University of California Regents.

Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11, 149–178. doi:10.1207/s15326977ea1103&4_2.

Bates, C., & Nettlebeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177–187. doi:10.1080/01443410020043878.

Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, 40(1), 23–38.

Beswick, J. F., Willms, J. D., & Sloat, E. A. (2005). A comparative study of teacher ratings of emergent literacy skills and student performance on a standardized measure. *Education*, 126(1), 116–138.

Björn, P., Aro, M., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learning Disability Quar*terly, 39(1), 58–66. doi:10.1177/0731948715594787.



- Catts, H. W., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, 48(3), 281–297.
- Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., ... Crouch, R. C. (2010). Selecting at-risk firstgrade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327–340.
- Davis, G. N., Lindo, E. J., & Compton, D. L. (2007). Children at risk for reading failure. Teaching Exceptional Children, 39(5), 32–37.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219–232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. The Journal of Special Education, 37, 184–192. doi:10.1177/00 224669030370030801.
- 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.
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. School Psychology Quarterly, 18, 52–65. doi:10.1521/scpq.18.1.52.20876.
- Fletcher, J., & Satz, P. (1984). Test-based versus teacher-based predictions of academic achievement. A three-year longitudinal follow-up. *Journal of Pediatric Psychology*, 9(2), 193–201. doi:10.1093/jpepsy/9.2.103.
- Flynn, J. M., & Rahbar, M. H. (1998). Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools*, 35(2), 163–172.
- Graney, S. B. (2008). General education teacher judgments of their low-performing students' short-term reading progress. *Psychology in the Schools*, 45(6), 537–549.
- Hintikka, S., Landerl, K., Aro, M., & Lyytinen, H. (2008). Training reading fluency: Is it important to practice reading aloud and is generalization possible? *Annals of Dyslexia*, 58, 59–79. doi:10.1007/s11881-008-0012-7.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. Review of Educational Research, 59, 297–313. doi:10.2307/1170184.
- Holopainen, L., Ahonen, T., & Lyytinen, H. (2001). Predicting reading delay in reading achievement in a highly transparent language. *Journal of Learning Disabilities*, 34(5), 401–414.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. School Psychology Review, 36(4), 582–600.
- Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research and Practice*, 24(4), 174–185.
- Kikas, E., Silinskas, G., & Soodla, P. (2015). The effects of children's reading skills and interest on teacher perceptions of children's skills and individualized support. *International Journal of Behavioral Development*, 39(5), 402–412.
- 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.
- 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, 265–296. doi:10.1023/A:1011117620895.
- Lerkkanen, M.-K. (2007). The beginning phases of reading literacy instruction in Finland. In P. Linnakylä & I. Arffman (Eds.), Finnish reading literacy. When quality and equity meet (pp. 155–174). Jyväskylä: University of Jyväskylä, Institute for Educational Research.

- Lerkkanen, M.-K., Ahonen, T., & Poikkeus, A.-M. (2011). The development of reading skills and motivation and identification of risk at school entry. In M. Veisson, E. Hujala, P. K. Smith, M. Waniganayake & E. Kikas (Eds.), Global perspectives in early childhood education: Diversity, challenges and possibilities (pp. 237–238). Frankfurt am Main: Peter Lang.
- Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, M., Siekkinen, M., & Nurmi, J.-E. (2006). The first steps study [Alkuportaat], ongoing. Finland: University of Jyväskylä.
- Lerkkanen, M.-K., Poikkeus, A.-M., & Ketonen, R. (2006). ARMI. Lukuja kirjoitustaidon arviointimateriaali 1. luokalle. [ARMI—a tool for assessing reading and writing skills in grade 1]. Helsinki: WSOY.
- Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. E. (2004). Reading performance and its developmental trajectories during the first and the second grade. *Learning and Instruction*, 14(2), 111–130. doi:10.1016/j.learninstruc.2004.01.006.
- Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006). Trajectories of reading development: a followup from birth to school age of children with and without risk for dyslexia. Merrill-Palmer Quarterly, 52(3), 514–546.
- Madelaine, A., & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52, 33–42. doi:10.1080/10349120500071886.
- Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools*, 48(4), 343–356. doi:10.1002/pits.20558.
- Mesmer, E. M., & Mesmer, H. A. E. (2008). Response to intervention (RTI): What teachers of reading need to know. *The Reading Teacher*, 62(4), 280–290. doi:10.1598/RT.62.4.1.
- Parrila, R., Aunola, K., Kirby, J. R., Leskinen, E., & Nurmi, J. E. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology*, 97(3), 299–319. doi:10.1037/0022-0663.973.299.
- Poskiparta, E., Niemi, P., & Lepola, J. (1994). Diagnostiset testit 1. Lukeminen ja kirjoittaminen. Turku: Turun yliopisto, Oppimistutkimuksen keskus. [Diagnostic Tests 1, Reading and Writing].
- Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppänen, P. H. T., Poikkeus, A.-M., & Lyytinen, H. (2007). Very early phonological and language skills: estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, 48(9), 923–931. doi:10.1111/j.1469-7610.2007.01763.x.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychol*ogy, 94(2), 143–174.
- Snowling, M. J., Duff, F., Petrou, A., & Schiffeldrin, J. (2011). Identification of children at risk of dyslexia: the validity of teacher judgments using 'Phonic Phases'. *Journal of Research in Reading*, 34(2), 157–170. doi:10.1111/j.1467-9817.2011.01492.x.
- Soodla, P., & Kikas, E. (2010). Teachers' judgment of students' reading difficulties and factors related to its accuracy. In A. Toomela (Ed.), Systemic Person-Oriented Study of Child Development in Early Primary School (pp. 73–94). Pieterlen: Peter Lang.
- 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. doi:10.1016/j. learninstruc.2015.02.002.
- Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary School Journal*, 111(4), 585–607.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819. doi:10.1002/pits.20113.



- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. doi:10.1037/a0077627
- Takala, M., & Ahl, A. (2014). Special education in Swedish and Finnish schools: Seeing the forest or the trees? *British Journal of Special Education*, 41(1), 59–81. doi:10.1111/1467-8578.12049.
- Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Lateemerging and resolving dyslexia: A follow-up study from age 3
- to 14. *Journal of Abnormal Child Psychology*, 43(7), 1389–1401. doi:10.1007/s10802-015-0003-1.
- Torppa, M., Poikkeus, A.-M., Laakso, M.-L., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter name knowledge and its relation to grade 1 reading achievement in children with and without familial risk for dyslexia. *Developmental Psychology*, 42(6), 1128–1142.
- Zumeta, R. O., Compton, D. L., & Fuchs, L. S. (2012). Using word identification fluency to monitor first-grade reading development. *Exceptional Children*, 78(2), 210–220.





II

SPECIAL EDUCATION TEACHERS' IDENTIFICATION OF STUDENTS' READING DIFFICULTIES IN GRADE 6

by

Riitta Virinkoski, Marja-Kristiina Lerkkanen, Kenneth Eklund & Mikko Aro 2020

Scandinavian Journal of Educational Research, 66(1), 59-72

https://doi.org/10.1080/00313831.2020.1833241

Reproduced with the kind permission of Taylor & Francis.



Scandinavian Journal of Educational Research





ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/csje20

Special Education Teachers' Identification of Students' Reading Difficulties in Grade 6

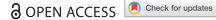
Riitta Virinkoski, Marja-Kristiina Lerkkanen, Kenneth Eklund & Mikko Aro

To cite this article: Riitta Virinkoski, Marja-Kristiina Lerkkanen, Kenneth Eklund & Mikko Aro (2020): Special Education Teachers' Identification of Students' Reading Difficulties in Grade 6, Scandinavian Journal of Educational Research, DOI: 10.1080/00313831.2020.1833241

To link to this article: https://doi.org/10.1080/00313831.2020.1833241

9	© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
	Published online: 27 Oct 2020.
	Submit your article to this journal 🗗
hh	Article views: 1369
Q [\]	View related articles ☑
CrossMark	View Crossmark data ☑
2	Citing articles: 2 View citing articles 🗗







Special Education Teachers' Identification of Students' Reading Difficulties in Grade 6

Riitta Virinkoski (10 a,b), Marja-Kristiina Lerkkanen (10 b), Kenneth Eklund (10 c) and Mikko Aro (10 d)

^aOpen University, University of Jyväskylä, Jyväskylä, Finland; ^bDepartment of Teacher Education, University of Jyväskylä, Jyväskylä, Finland; ^cFaculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland; ^dDepartment of Education, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

This study investigated special educational needs (SEN) teachers' (n = 29) assessment practices and the accuracy of their ratings of the students' (M age = 12.75 years, n = 55) skill levels in reading fluency and reading comprehension. Teachers rated their sixth grade students' fluency and comprehension on a three-point scale, and the students were also tested in group tests. Results showed that SEN teachers used several assessment practices simultaneously but mostly relied on observations. The correlations between the teacher ratings and the test scores were significant but moderate in fluency and weak in comprehension. Only two thirds of low-performing students having difficulties in fluency or comprehension were identified. Additionally, identification of students with typical reading comprehension was inaccurate.

ARTICLE HISTORY

Received 3 September 2019 Accepted 12 August 2020

KEYWORDS

Reading difficulties; reading fluency; reading comprehension; special education teachers; sensitivity; specificity

Teachers' student evaluations and perceptions provide valuable information for individual support or instructional decisions throughout their students' educational paths (Bailey & Drummond, 2006; Feinberg & Shapiro, 2009). Teachers' assessments of students' reading skills might also have longterm effects for their further reading development and future academic opportunities (Paleczek et al., 2017). However, studies regarding oral reading fluency (Begeny et al., 2008; Feinberg & Shapiro, 2009) have shown that teachers might find it difficult to accurately rate their students' reading levels as low, average, or high. Teachers seem to be more capable of comparing an individual's performance to that of the student's peers than evaluating the student based on his or her actual test scores (Begeny et al., 2011).

In addition to classroom teachers, special educational needs (SEN) teachers and remedial reading teachers (Ise et al., 2011) play an essential role in evaluating students' needs for reading support, identifying literacy difficulties, and providing individualized targeted support. In Finland, part-time special education for students with learning difficulties in mainstream settings is used extensively, providing low-threshold support for any student, and usually, the SEN teachers work in one school only. No formal diagnosis is required to receive learning support in Finnish schools (Björn et al., 2016).

This study focuses on Finnish SEN teachers and students receiving part-time special education in the sixth grade. SEN teacher training in Finland is a master's degree program, comprising theory and practice related to individual and small-group instruction, the application of various assessment tools, support in reading, writing, mathematics, and communication, as well as behavioral and socio-emotional challenges (Takala et al., 2015). Students with reading difficulties (RD) receive



remedial instruction during or after school by classroom teacher, part-time special education by SEN teacher, or co-teaching by classroom teacher and SEN teacher during literacy lessons (Lerkkanen, 2007). Part-time special education is typically provided one to two hours per week in small groups of three or four students (Holopainen et al., 2017).

Prior studies have shown that in transparent orthographies, such as Finnish, students with RD usually experience persistent problems with reading fluency rather than accuracy (Aro & Wimmer, 2003; Eklund et al., 2015; Soodla et al., 2019). Some students' RD may not manifest themselves in the early school years but might emerge later and become persistent (Kent et al., 2018; Torppa et al., 2015; Wanzek et al., 2010). Recent findings among second through tenth grade students (Catts et al., 2012; Solis et al., 2014; Torppa et al., 2015) have confirmed that many students will continue to demonstrate RD after the early school grades. In the middle and secondary grades, the role of written texts as primary repositories for learning academic content and acquiring knowledge is emphasized (Oslund et al., 2018).

Although identification of students' RD during primary school has been widely studied, research on teachers' assessment of RD and the accuracy of the assessments before transition to lower secondary school (i.e., at the end of Grade 6 in the Finnish educational system) is still restricted. Particularly current research on the accuracy of SEN teachers' assessment practices is lacking, and this study adds to prior findings of Virinkoski et al. (2018) concerning teachers' assessments of prereading skills in Grade 1.

Teachers' Assessment Practices

Decision making in education as well as planning instruction and support require collecting data and monitoring student progress (Cornelius, 2013). Assessments of student learning are usually done in two ways, depending on the objective. Teachers use formative (or informal) assessments to collect data on student's current skills or to improve student's learning by implementing more personalized instruction. However, teachers employ summative (or formal) assessments to assess student's knowledge after completing a certain learning sequence (Cornelius, 2013; Dixson & Worrell, 2016; Dolin et al., 2018). One can also define assessment tools as high-stakes or low-stakes tools. The former is connected to the final assessment of how much the student has learned at a certain assessment point, and they usually take place when an instructional segment (e.g., semester) has ended, such as statewide or national tests (Dixson & Worrell, 2016; Dolin et al., 2018). The latter refers to low-stakes continuous instruction, comprising assessments before and during instruction, typically in the form of observations, self-evaluations, and curriculum-based measures (CBMs; Dixson & Worrell, 2016; Dolin et al., 2018). Generally, teachers' assessment practices can be divided into tests (screening or individual tests), CBMs, and qualitative assessments, such as observations and checklists (Bailey & Drummond, 2006; Südkamp et al., 2012).

Teachers can apply either direct or indirect measures when assessing students' reading skills (Begeny et al., 2011; Woolley, 2008). Direct measures refer to tests, which usually have distinctive limitations and strengths. For example, accuracy of screening measures differs in terms of sensitivity and specificity (Barth et al., 2014; Catts et al., 2015; Compton et al., 2010). Woolley (2008) found that teachers who relied solely on indirect measures, such as classroom observations, often made inaccurate assessments of students' actual reading performances, tended to overestimate students' abilities, and judged high-performing readers better than low- or average-performing readers. Feinberg and Shapiro (2009) also discovered that the accuracy of teachers' ratings of reading fluency and comprehension through observation was low in comparison to the identification of low student reading performance using CBMs and standardized achievement tests. Their study also showed that teacher reports may not produce specific information compared to some other academic measures, such as CBMs.

CBMs are one way to assess students' literacy progress toward long-term curriculum goals, which are also the main tools in the response to intervention (RTI) framework (Marchand & Furrer,

2014; Stecker et al., 2005) for recognizing learning difficulties and the risk for RD. CBMs can be used in general, remedial, and special education to monitor students' progress in, for example, overall school performance, or to screen students at specific time points to determine their level of risk for academic failure (Reschly et al., 2009; Zumeta et al., 2012). Prior studies have shown that using CBMs in conjunction with standardized procedures to track students' reading development can lead to higher identification accuracy of struggling readers as well as improvements in reading achievement (Deno, 2003; Feinberg & Shapiro, 2009; Wayman et al., 2007). Some studies (e.g., Woolley, 2008) have also indicated that teacher-designed instruments are more informative yet less reliable because their content, assessment conditions, and assessor variables may differ. However, by combining various assessment instruments and observation tools, the quality of teachers' assessments can be improved (Feinberg & Shapiro, 2009; Woolley, 2008).

Assessment of Reading Fluency and Comprehension

To gain a comprehensive understanding of a student's reading skills, especially when he or she has RD, it is necessary to assess both reading fluency and reading comprehension. One key factor in learning to read fluently is automatic word recognition, which develops through consistent practice, repetition, and reading a wide range of various texts (Kuhn et al., 2010). Fluent readers are able to identify words in the text without conscious effort because their decoding has become automatic (Meisinger et al., 2010; Wagner & Espin, 2015). Many researchers (e.g., Hudson et al., 2008; Kuhn et al., 2010; Meisinger et al., 2010) agree that reading fluency comprises decoding accuracy and automaticity, both of which are connected to reading comprehension (e.g., Kim et al., 2015; Leppänen et al., 2008). The more accurate and automatic the decoding is, the more the readers' resources can be invested in reading comprehension (Leppänen et al., 2008). One common definition of reading fluency includes reading accuracy and rate, and reading fluency is usually operationalized as the number of correctly read items within a time limit (Hudson et al., 2008; Kuhn et al., 2010).

The goal of reading instruction is comprehension, which is a complex ability, covering "construction of meaning through contact and involvement with written language" (Cadime et al., 2017, p. 592). Reading comprehension requires word recognition (i.e., decoding), language comprehension, inference making, and information integration (Cain & Oakhill, 2006; Kuhn et al., 2010; Mancilla-Martinez & Lesaux, 2010). Some examples of common measures to assess reading comprehension are silent reading tasks (e.g., Oslund et al., 2018), either written or oral retell tasks (Bernfeld et al., 2013; Reed & Vaughn, 2012), and picture matching tasks (García & Cain, 2014).

Associations between Teacher Ratings and Test Scores

A high-quality screening measure should be able to accurately identify those students who are having RD as well as those whose development is proceeding according to expectations (Ball & O'Connor, 2016). Of these two aspects, the sensitivity of the assessment tool (i.e., its accuracy in identifying students with problems) has usually been considered important so that support can be allocated to those students who need it most. In contrast, specificity refers to the accuracy of an assessment tool to correctly identify students who are not at risk (Ball & O'Connor, 2016). An acceptable level of classification accuracy for sensitivity the percentage is considered to be 90% or above, and for specificity, it should be at least 80% (Catts et al., 2015; Compton et al., 2010). A high sensitivity level is often accompanied by a low specificity level (Ball & O'Connor, 2016; Barth et al., 2014). However, accurate identification would require a high percentage of true positives, whereas the number of false positives should remain manageable (Compton et al., 2010). Prior studies investigating the accuracy of teacher ratings as compared to test scores (Begeny et al., 2008; Soodla & Kikas, 2010) have shown that the assessment of typically performing students is more accurate than of low-performing students.

Prior studies have also demonstrated that despite relatively high overall correlations between teachers' ratings and students' actual test scores, teachers may systemically over- or underestimate students' performances (Feinberg & Shapiro, 2009; Martin & Shapiro, 2011; Martínez et al., 2009). For instance, teachers might base their decision for allocating support on the student's former weaknesses in reading or previous identification of the student as requiring special educational support (Soodla & Kikas, 2010). Additionally, students with low academic performance are usually judged less accurately than typically performing students (Feinberg & Shapiro, 2009; Kikas et al., 2017; Soodla & Kikas, 2010). According to a meta-analysis by Südkamp et al. (2012), the more specific subskills of reading (i.e., reading comprehension as opposed to overall reading performance) teachers were asked to evaluate, the more congruent with the test scores the judgments were. Nevertheless, in a study by Karing and Artelt (2013) teachers were more accurate in assessing students' general abilities (e.g., reading skill) than specific ones.

Feinberg and Shapiro (2009) studied teachers' judgment accuracy of second to fifth grade students' reading skills, and the correlation in reading fluency was moderate (.47), and strong in reading comprehension (.60), and a study by Begeny et al. (2008) reported correlations ranging from .53 to .79 in oral reading fluency (1st to 3rd grade students). Further, Paleczek et al. (2017) reported in their study rather strong correlations (in decoding .57, and .69 in reading comprehension) between teacher judgments and test scores, at the end of Grade 3. Also Kikas et al. (2017) studied 3rd grade students' reading comprehension, and a rather strong correlation (.55) was reported between teacher judgments and test scores.

The Aim of the Present Study

The aim of this study is to investigate SEN teachers' assessment practices and the accuracy of their ratings of reading fluency and reading comprehension in Grade 6, before transition to lower secondary school. The research questions are as follows:

- (1) What kinds of assessment practices do SEN teachers use to evaluate students' reading performances, and how do they rank different practices? Based on prior findings, we hypothesize that the most often-applied assessment tools are observations, CBMs, and achievement tests (Bailey & Drummond, 2006; Südkamp et al., 2012).
- (2) To what extent are SEN teachers' ratings of sixth-grade students' reading fluency and reading comprehension skills associated with students' test scores for the same skills? We expect moderate correlations between teachers' ratings and test scores of reading fluency and reading comprehension, since the correlations in prior studies have mostly varied from rather strong (Paleczek et al., 2017) to moderate (Feinberg & Shapiro, 2009).
- (3) How accurate are SEN teachers' perceptions of their students' reading fluency and reading comprehension skill levels (low performing or typically performing), compared to the students' test scores? We expect that the sensitivity and specificity rates do not reach the acceptable levels (90%, and 80%, respectively), and that there are no major differences between the accuracy of teachers' ratings of both reading fluency and reading comprehension as compared to test scores (Feinberg & Shapiro, 2009; Paleczek et al., 2017). We also anticipate that teacher ratings of typically performing students are more accurate as compared to those of the low-performing students (Begeny et al., 2008; Coladarci, 1986; Paleczek et al., 2017).

Method

Participants and Procedure

The data were drawn from a follow-up study of 1,880 children (Lerkkanen et al., 2006) from four municipalities in Finland comprising one whole age cohort of children in three medium-sized

towns and half of an age cohort in one municipality. Parental education levels in the data set are close to the Finnish national average (Eurostat, 2013). The childrens' caretakers were asked for written consent at the beginning of the study.

Teacher sample included 29 (90% female) SEN teachers (M age 41.41 years, SD = 9.99 years), 65% working in one school, and 35% working in several schools. Most teachers' basic education was a master's degree from a classroom teacher program combined with a SEN teacher qualification (52%) or a master's degree from a SEN teacher program (45%). Three percent had another basic education such as a BA degree as a kindergarten teacher combined with a SEN teacher qualification. Twenty-eight percent had 1-5 years of professional experience, 31% 6-10 years, 17% 11-15 years, and 24% more than 15 years.

Student sample (n = 55, 65% boys; M age = 12.75 years, SD = 0.39) was selected from a more intensively followed sub-sample of 598 students drawn from the whole sample of 1,880 students including both students identified as being at-risk for RD (n = 277) and control children who were not at-risk for RD (n = 321). The selection criteria were that the children had been followed up individually since kindergarten, as determined by the researchers at the end of kindergarten year, on the basis of four criteria related to risk for RD: children's initial phoneme identification skill, letter knowledge, rapid automatized naming, and parental reports of their own RD (see Lerkkanen et al., 2011). The risk for RD was determined at the end of the kindergarten year based on three tests (letter knowledge, phonemic awareness and rapid automatized naming) and parents' self-reported RD (mothers or fathers indicated on a questionnaire that they had mild or severe problems in reading at school age) (Lerkkanen et al., 2011). These variables were suggested by metaanalyses and familial dyslexia follow-up studies (e.g., Lyytinen et al., 2006). Children were identified as at-risk for RD if they scored at or below the 15th percentile in at least two of the measured skill areas or if they scored at or below the 15th percentile in one skill area, and parental questionnaire indicated family risk (see Lerkkanen et al., 2011).

Data collection for this study was carried out during the spring term of Grade 6. The SEN teachers were sent a list of the students that were followed more intensively in their schools but they did not know which group (at-risk for RD or control group) the children belonged to, and they were asked to rate the ones (1-6 students) that had received part-time special education in Grade 6. Reasons as well as the amount and form of part-time special education were reported by the teachers, and students' reading fluency and reading comprehension were assessed with three groupadministered tests by trained testers. Altogether, 23 out of 29 teachers taught sixth-grade students who met the selection criteria, and they also returned the student evaluation forms (n = 55).

Measures

SEN Teachers' Assessment Practices of Reading

Teachers reported their assessment practices on the questionnaires by answering the question, "In your opinion, what kinds of practices regarding the assessment and follow-up of reading and writing skills are most important in your work at the moment?" (Items were rated from first to sixth most important or not in use). The items were: exams, own observations, assessment forms, tests, discussions/interviews (later discussions), and something else (specify). Teachers were also asked to give additional information about some evaluation practices (e.g., tests, and discussions) by responding to open-ended questions about, for example, the kinds of tests they used and with whom they had discussions about the students' reading performances.

Ratings of Students' Reading Skills by SEN Teachers

The data of this study comprised teachers' ratings of students' reading fluency and reading comprehension skills using a three-point rating scale: clear problems, mild problems, and no problems. Based on the rather small student sample (n = 55) and the focus of the study being on the teachers' perceptions of students' RD in general (clear or mild problems), the categories "clear problems" and "mild problems" were pooled together as follows: 0 = no problems, and 1 = clear or mild problems. On teachers' questionnaire, teachers were also asked how often they were currently teaching the students (this information was available concerning 52 students). According to the responses, 51% of the 6^{th} students were provided special education regularly (1–2 times a week), and 40% irregularly (some hours a month), and 9% periodically.

Reading Fluency and Comprehension

Students' reading fluency was tested using two group-administered tests: a word reading fluency test and a sentence reading fluency test. Word reading fluency task was a subtest of a standardized Finnish reading test battery for primary schools (ALLU; Lindeman, 1998). Each of the 80 items consisted of a picture with four phonologically similar words next to it. Words and pictures were frequently used words familiar to children and they were instructed to read the four words silently and connect the picture with the correct, semantically matching word by drawing a line. The score was the number of correct answers within a two-minute time limit (maximum = 80). Pearson correlation coefficients between subsequent time points was .62 (Grades 4 and 6). Sentence reading fluency task was a Finnish adaptation of the Salzburger Lese-Screening test (SLS; Mayringer & Wimmer, 2003), Luksu, which is similar to the Woodcock-Johnson sentence verification task (Woodcock et al., 2001). Each student was instructed to read a sentence as quickly as he/she could and decide whether the given statement (e.g., "Blueberries are yellow") was true or false. The score was the number of correct responses given within a two-minute time limit (maximum = 60). Between Grades 4 and 6, the Luksu test correlation was .68. Fluency score was the mean of the standardized scores of the two tasks, for which the Cronbach's alpha was .64.

A group-administered subtest of a standardized Finnish reading test for primary school (ALLU; Lindeman, 1998) was used to assess reading comprehension. Students were instructed to read silently a short story and then answer 12 multiple-choice questions based on the text, each with four alternatives. A point was scored for each correct answer (maximum = 12). Students completed the task at their own pace, but the maximum time allotted was 45 min. Cronbach's alpha was .66 in Grade 6.

Data Analyses

IBM SPSS Statistics 24 program was used to obtain descriptive statistics and to perform the analyses. Spearman's rank correlation coefficient was used to calculate the correlations between the teachers' ratings and the test scores. A one-tailed test was conducted based on the assumption that both teachers' ratings and test scores were parallel. Sensitivity and specificity rates of the teachers' ratings were also calculated in order to show the accuracy of the ratings. For statistical testing of RQ 3, binary logistic regression analyses were conducted. Students' dichotomized test scores were used as dependent variables and teachers' dichotomized ratings as independent variables. Reading fluency and reading comprehension test scores were dichotomized to low-performing students and typically performing readers using the 16^{th} percentile (-1.0 SD), based on the large First Steps follow-up sample (N = 1,880) as the cut-off score. To be classified as a low-performing student in reading fluency the student had to score below the 16^{th} percentile in the mean of the standardized scores of two reading fluency tasks and regarding the comprehension task, the criterion for the low-performing student was maximum four correct responses out of 12.

Results

The first research question was what kinds of assessment practices SEN teachers use to evaluate their students' reading skills. Five given items for assessment practices were own observations, discussions, tests, assessment forms, and exams. The results first showed that all teachers used several practices. Further, according to the responses, 66% (n = 19) of teachers used five different practices,

	Rank Mean	SD	Assessment practice not in use (n)
Own observations	1.89	1.29	1
Discussions	2.55	1.40	0
Tests	2.74	1.23	2
Assessment forms	3.13	1.36	6
Exams	3.35	1.16	3

Note. Rank means were calculated from the SEN teachers' assessment practices ratings (1 = the most important to 5 = the least important).

24% used four practices (n = 7), and 10% used three practices (n = 3). Means and standard deviations of the assessment practices reported on the teachers' questionnaires are presented in Table 1.

Teachers had arranged the assessment practices with respect to importance, using a six-point scale: 1 = the most important, 2 = the second most important, etc. Because there were only a few responses for choice six, responses were recoded so that the least important choices (i.e., fifth most important and sixth most important) were combined into one category (5 = the least important). Additionally, in cases where teachers did not use a certain assessment practice, they were asked to leave the choice in question empty.

According to the responses, two kinds of indirect assessment practices were the most important of all the given items. The first was own observations, which was ranked the most important or the second most important practice by 71% of the teachers (see Table 2). Another indirect practice, discussions, was ranked the second most important assessment practice, with 56% of the teachers listing it as the most or the second most important practice used. Teachers reported that they had discussions with the parents, other teachers, usually the classroom teachers, and the students themselves. Direct assessment practices (i.e., tests) were ranked as the most important by only 15% (n = 4) of the teachers, and 7% (n = 2) reported not using tests at all. Usually, tests were word reading fluency or silent reading comprehension tests or e.g., tests where students had to differentiate words from longer chains of words. Assessment forms were, for example, materials connected to reading achievement tests and less than 9% of the teachers listed this as the most important practice.

Next, we studied the associations between the teachers' ratings and the sixth-grade students' test scores. Teachers had rated the students' fluency and comprehension skills using three-point rating scales, and three tests were used to evaluate the same skills. Table 3 shows the descriptive statistics of all the measures in the sample regarding research questions 2 and 3.

All of the students' test scores were normally distributed, whereas the teachers' ratings were left-skewed. Therefore, Spearman's correlations were used when examining the associations between the test scores and teacher ratings. The results first showed that teachers considered most of the students as having no problems with fluency. Second, the mean of the teachers' ratings was

Table 2. The Importance of Assessment Practices to SEN Teachers.

Assessment practice	The most important n (%)	The 2 nd most important n (%)	The 3rd most important n (%)	The 4 th most important n (%)	The least important n (%)	Total n (%)
Own observations	17 (60.72)	3 (10.71)	3 (10.71)	4 (14.29)	1 (3.57)	28 (100)
Discussions	8 (27.59)	8 (27.59)	7 (24.14)	1 (3.44)	5 (17.24)	29 (100)
Tests	4 (14.81)	11 (40.75)	1 (3.70)	10 (37.04)	1 (3.70)	27 (100)
Assessment forms	2 (8.70)	7 (30.42)	6 (26.09)	2 (8.70)	6 (26.09)	23 (100)
Exams	2 (7.69)	3 (11.54)	10 (38.46)	6 (23.08)	5 (19.23)	26 (100)

Note. n = number of responses; number of SEN teachers varied between 23 and 29, depending on whether they reported using the practice or not.



Table 3. Descriptive Statistics of the Measures in This Study: Teachers' Ratings and Students' Scores for Reading Tasks.

	N	M	SD	Range
Fluency: SEN teachers' ratings	55	2.49	0.69	1–3
Comprehension: SEN teachers' ratings	54	1.87	0.75	1–3
Word recognition fluency task	55	42.20	10.13	18-63
Sentence fluency task	55	25.42	6.63	10-41
Reading comprehension task	55	5.20	2.76	1–12

Note. N = number of students; M = mean of ratings/scores; rating scale: 1 = clear problems, 2 = mild problems, 3 = no problems.

Table 4. Identification Accuracy of SEN Teachers' Ratings' Merged Categories 1 and 2 (Clear and Mild Problems) and Students' Test Scores (16th percentile).

Reading skill	Sensitivity (%)	Specificity (%)	TP % (n)	FP % (n)	TN % (n)	FN % (n)
Reading fluency $(n = 55)$	63	69	18 (10)	22 (12)	49 (27)	11 (6)
Reading comprehension $(n = 54)$	70	20	13 (7)	64 (35)	17 (9)	6 (3)

Note. TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative. Sensitivity = TP/(TP + FN); specificity = TN/ (TN + FP).

remarkably lower for comprehension, meaning that based on the teachers' ratings, students had more difficulties in comprehension than in fluency.

Investigation of the associations between the teachers' ratings for fluency and comprehension as well as the students' test scores showed that teachers' ratings for reading fluency were significantly correlated (.39, p < .01) to students' performances in the two fluency tasks and there was also a significant correlation between the teachers' ratings and the reading comprehension test scores (.24, p < .05).

Finally, we analyzed the accuracy of the teachers' perceptions of students' skill levels: how accurately they were able to separate low achievers in fluency and comprehension (sensitivity), from those who, according to the test scores, were typically performing readers (specificity). Students were classified into two groups (i.e., low achievers and typically performing readers) based on their test scores, using the 16^{th} percentile value of a large population-based sample (N = 1,880) of the First Steps follow-up study as the cut-off score. Regarding teachers' ratings, students in the "clear problems" and "mild problems" categories were considered low achievers, whereas students in the "no problems" category were classified as typically performing readers.

Table 4 shows sensitivity and specificity rates as well as true positives, false positives, true negatives, and false negatives concerning low-performing students in fluency and comprehension. Altogether, 16 students were classified as belonging to the low-performing group based on their fluency scores (mean of the two fluency z-scores), whereas teachers had rated a total of 22 students as low achievers in fluency. Likewise, for comprehension, 10 and 42 low-performing students were identified (based on test scores and teachers' ratings, respectively).

According to logistic regression analyses (see Table 5), teachers' ratings were significantly associated with students' categorical reading fluency test scores (χ 2 (1) = 4.72, p = .030). Instead, teachers' ratings were not associated with students' categorical reading comprehension test scores ($\chi 2$ (1) =

Table 5. Two Skill-Specific Logistic Regression Models for Predicting Reading Fluency and Reading Comprehension Performance in Grade 6.

SEN teachers' ratings	В	S.E.	OR	р	R ²
Reading fluency	1.322	.622	3.750	.034	.117
Reading comprehension	– .511	.785	.600	.515	.012

Note. B = value for logistic regression equation for predicting dependent variable from independent variable; S.E. = Standard Error; OR = odds ratio; $R^2 = adjusted version of the Cox & Snell R-square.$

0.41, p = .524). Teachers had identified difficulties in comprehension more frequently (78%) than in fluency (40%), even though both figures included a prominent number of false positives (55% and 83%, in fluency and comprehension, respectively). Sensitivity rate was rather low and below the acceptable rate (see Compton et al., 2010; Johnson et al., 2010) for both fluency (63%) and comprehension (70%). Additionally, specificity rates for fluency did not quite reach an acceptable accuracy level although it was rather high (69%), and for comprehension, the very low (20%) specificity rate revealed difficulties with identifying typically performing readers (see e.g., Johnson et al., 2010).

Discussion

Aims of this study were to examine SEN teachers' assessment practices of reading, to measure the accuracy of their ratings concerning sixth-grade students' reading fluency and reading comprehension compared to the students' test scores, and to analyze how accurately teachers identified lowperforming and typically performing students.

First, partly as we expected, the findings showed that the most important assessment practices for SEN teachers were qualitative, such as observations and discussions. This finding is also supported by prior research (Virinkoski et al., 2018), which indicated that most classroom teachers, but also SEN teachers, relied heavily on qualitative practices to identify students at risk for RD in Grade 1. Opposite to what would have been anticipated (see Feinberg & Shapiro, 2009; Wayman et al., 2007), achievement tests were not among the most important assessment practices, although they were widely used (93%) by the teachers, together with some other tools. One explanation may be that in Finland, a student's poor test performance on a standardized test is not the sole reason for providing part-time special education. Instead, SEN teachers use their perceptions and pedagogical knowledge in deciding whether to provide support. Nevertheless, in this study, all teachers used several assessment practices in parallel, but they mainly preferred indirect, qualitative assessment practices, such as observations and discussions with students, teacher colleagues, or parents, compared to CBMs or test evaluations. According to prior studies (Feinberg & Shapiro, 2009; Woolley, 2008), assessments based solely on observation are often inaccurate; instead, using various assessment practices together can improve the accuracy, especially when the standardized procedures are combined with CBMs (Feinberg & Shapiro, 2009; Wayman et al., 2007). As in this study, using mainly qualitative practices can make the teachers' assessments less reliable because of assessor variables and the conditions of the assessment, for example (see Martínez et al., 2009).

Second, we hypothesized that the correlations between the teachers' ratings and the test scores would be moderate. However, only the correlation for reading fluency was moderate (.39), but the correlation for reading comprehension was weak (.24). Further, logistic regression analyses confirmed a significant association between the teachers' ratings and the fluency test scores, unlike those of comprehension. The moderate and weak correlations between the teachers' ratings and the test scores found in this study are substantially lower, compared to those reported in a number of former studies. For instance, Paleczek et al. (2015), and Paleczek et al. (2017) reported significant correlations of .60 and .57, respectively, for decoding, and .60 to .69, respectively, for reading comprehension. In addition, a study by Kikas et al. (2017) reported a correlation of .55 for reading comprehension. In the first two studies, standardized test scores (individual and group-administered tests) were compared to the teacher ratings using a 4-point Likert scale, and in the study by Kikas et al. (2017), the scores of a test measuring academic skills were compared to the 5-point rating scale the teachers used.

One explanation for the weak and moderate correlations in this study may due to teachers using a 3-point rating scale while reading tests had continuous scales. Only a thin line may have existed between the classifications of those students who were close to the cut-off score (see also Branum-Martin et al., 2012). Prior studies have also presented reasons for poor associations between teacher ratings and test scores, such as teachers using various assessment methods (e.g., direct or indirect) and the way the data were analyzed (Begeny et al., 2008). Inconsistencies between the teachers' ratings and the test scores in the present study may also reflect the nature of SEN teachers' work in

later primary school grades. Concerning this study, teachers may not have had the opportunity to gain adequate knowledge about their students, due to infrequent or even periodic teaching (almost 50% of the teachers) or limited contact with some students.

Finally, according to what we expected, teachers' judgments of both reading fluency and reading comprehension were quite inaccurate, compared to the test scores. There was only a minor difference between the sensitivity rates of reading fluency and reading comprehension in favor of reading comprehension (63% and 70%, respectively). These findings indicate that at least 30% of the sixth-grade students struggling with RD were unidentified. Our findings are in line with prior studies, indicating that teachers' judgments of low-performing students in reading have been inaccurate, and that teachers tend to overestimate the skills of low-performing students (Begeny et al., 2008; Feinberg & Shapiro, 2009). For instance, in a study by Soodla and Kikas (2010), 33% of low-performing students in reading comprehension had been correctly judged. However, assessing sixth-grade students' reading fluency can be quite challenging for SEN teachers for several reasons. First, they might not get many opportunities to observe students' reading aloud or their reading fluency; as mentioned earlier, some students were taught rather irregularly by the SEN teachers. Additionally, even though students may have received part-time special education for RD in the earlier school grades, in Grade 6, some students had received support for mathematics difficulties, for example, instead of RD. Second, decisionmaking based on test scores without explicit cut-off scores for distinguishing low-performing students from those performing typically can be difficult. Third, another explanation for the moderate sensitivity rate could also be due to the students' unexpected poor test performances in group-administered test situation, for example, not measuring fluency but possibly students' concentration and attention, as well as their level of executive functioning.

We expected that teachers' ratings would be more accurate for typically performing students than low-performing students (Coladarci, 1986; Paleczek et al., 2017). As noted earlier, according to Compton et al. (2010), for accurate identification of students with typical performance, specificity rate should be at least 80%. In this study, our findings were rather contradictory. First, specificity rate of reading fluency was rather high (70%) but below the optimal rate, indicating that 30% of the typically performing students were unidentified. Second, concerning reading comprehension, most students were incorrectly identified as low achievers (specificity 20%), even though their test scores indicated typical performance. This finding is remarkably lower than presented in a study by Soodla and Kikas (2010), where the accuracy of teachers' judgments of typically performing students in reading comprehension was 92%. Correlations obtained in this study were attenuated by poor reliability in the measures used (fluency .64 and comprehension .66). Similarly, teachers' judgments of students by category of reading ability appeared fairly inaccurate but possibly teachers' perceptions of "mild or clear problems" did not align well with the 16th percentile used on the test scores. Thus, applying lower or higher cut-score may have resulted to a different classification accuracy.

In Finland, provision of part-time special education to struggling students by SEN teachers has been an efficient means to narrow the gap between high and low achievers, and its emphasis has been on prevention (Itkonen & Jahnukainen, 2010). However, as the present study reveals, the only nationally standardized test for Finnish SEN teachers in primary schools is currently the ALLU test (Lindeman, 1998). This study underlines that teachers need reliable assessment tools throughout primary grades to monitor students' reading progress systematically and continuously. For instance, as our study also shows, support decisions based mainly on teachers' own perceptions and observations of students' performances can lead to inaccurate assessments (see also Soodla et al., 2019). One solution for better judgment accuracy could be a structured assessment tool designed for special education purposes enabling teachers to rank-order students' reading performance and compare the rankings with the reading test scores.

It has also been indicated that although some students do not show any difficulties in their early reading skills, they might turn up during their later school years (Kent et al., 2018; Torppa et al., 2015; Wanzek et al., 2010). Thus, we suggest that SEN teacher's role in supporting students' literacy learning is worth further investigation (see also Soodla et al., 2019). Areas to explore are how



students' skills should be assessed to gain higher sensitivity and specificity, kinds of assessment and support practices SEN teachers use with students who have RD, and what the most effective practices are and why.

Limitations

Despite its contributions to current research, the present study has some limitations that future studies should address. First, sample sizes of both SEN teachers and the sixth-grade students were rather small. This might lower the reliability and generalizability of the results. It also restricted our possibilities to investigate the background variables and their potential effects on the accuracy of the assessments. In addition, if the data had enabled us to study the mediating factors (frequency of support, teachers' work experience etc.), we might have gained a deeper understanding of this study's findings. Unfortunately, inadequacies connected to these factors in the research data made this kind of investigation impossible.

Second, in this study, students' test scores from three group-administered achievement tests constituted the basis for our analyses and students' performance levels (low performance and typical performance) were defined using the cut-off score of 16th percentile in the reading tests. However, using of three teacher ratings' categories "clear problems", "mild problems" and "no problems" separately in the analyses, especially the sensitivity percent decreased significantly, probably due to the fact that the three-point scale was not continuous in nature, so the distances between the three categories were not equal. Moreover, observations in each category were rather small, which did not enable using this measure as a nominal scale measure. Due to these reasons, we chose to dichotomize the options (0 = no problems, 1 = mild or clear problems), and focus on teachers' perceptions of students having RD or not. Third, the research data did not enable a closer investigation of the teachers' qualitative evaluations (e.g., their own observations), which proved the most important practice. Therefore, further research is needed about how these evaluations are conducted, what kinds of data collection modes are used in observations, and which of them are the most practical and efficient for accurate identification of students' RD.

Conclusions

The aim of this study was to fill the current gap in the existing reading assessment research by showing that in Finland's present educational system, SEN teachers play an essential role in evaluating, identifying, and supporting students who struggle with RD throughout the elementary grades. Based on our findings, the correspondence between SEN teachers' ratings and the test scores was not strong, which indicates the need for developing and deploying more systematic assessment tools for RD. They should also be applicable for identifying and monitoring the progress of upper elementary grade students' RD. Sixth-grade students' rather low sensitivity rates are especially alarming, which indicate that approximately 30% of students with poor reading test performance remain unidentified with current assessment practices, resulting in inadequate support for RD. It is hoped that in the future, Finnish teachers will have a range of standardized reading assessment tools to support their practices in reading interventions.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This study has been financed by the Academy of Finland (No. 292466 for 2015–2019; No. 268586 for 2013–2017) and a grant from the Ellen and Artturi Nyyssönen Foundation (2017, No. 03042017).



ORCID

Riitta Virinkoski http://orcid.org/0000-0002-7270-8230

Marja-Kristiina Lerkkanen http://orcid.org/0000-0002-5709-5800

Kenneth Eklund http://orcid.org/0000-0003-3416-4700

Mikko Aro http://orcid.org/0000-0002-0545-0591

References

- 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/10.017.S0142716403000316
- Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11(3), 149–178. https://doi.org/10.1207/s15326977ea1103&4_2
- Ball, C. R., & O'Connor, E. (2016). Predictive utility and classification accuracy of oral reading fluency and the measures of academic progress for the Wisconsin knowledge and Concepts Exam. Assessment for Effective Intervention, 41(4), 195–208. https://doi.org/10.1177/1534508415620107
- Barth, A. E., Stuebing, K. K., Fletcher, J. M., Denton, C. A., Vaughn, S., & Francis, D. (2014). The effect of reading duration on the reliability and validity of middle school students' ORF performance. *Assessment for Effective Intervention*, 40(1), 53–64. https://doi.org/10.1177/1534508414545643
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43–55. https://doi.org/10.1037/1045-3830.23.1.43
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, 40(1), 23–38. https://doi.org/10.1080/02796015.2011.12087726
- Bernfeld, L. E. S., Morrison, T. G., Sudweeks, R. R., & Wilcox, B. (2013). Examining reliability of reading comprehension ratings of fifth grade students' oral retellings. *Literacy Research and Instruction*, 52(1), 65–86. https://doi.org/10.1080/19388071.2012.702187
- Björn, P., Aro, M. T., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learning Disability Quarterly*, 39(1), 58–66. https://doi.org/10. 1177/0731948715594787
- Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2012). 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
- Cadime, I., Rodrigues, B., Santos, S., Viana, F. L., Chaves-Sousa, S., do Céu Cosme, M., & 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
- Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76(4), 697–708. https://doi.org/10.1348/000709905X69807
- Catts, H. W., Compton, D. L., 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., Nielsen, D. C., Bridges, M. S., Liu, Y. S., & Bontempo, D. E. (2015). Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*, 48(3), 281–297. https://doi.org/10.1177/0022219413498115
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141–146. https://doi.org/10.1037/0022-0663.78.2.141
- Compton, D. L., Fuchs, L., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, 102(2), 327–340. https://doi.org/10.1037/ a0018448
- Cornelius, K. E. (2013). Formative assessment made easy: Templates for collecting daily data in inclusive classrooms. *Teaching Exceptional Children*, 45(5), 14–21. https://doi.org/10.1177/004005991304500502
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, 37(3), 184–192. https://doi.org/10.1177/0022466903037003080 1
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, 55(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989
- Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. In J. Dolin & R. Evans (Eds.), *Transforming assessment. Through an interplay between practice, research and Policy* (pp. 53–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-63248-3.



- 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. https://doi.org/10.1037/a0037121
- Eurostat regional yearbook 2013: Education. (2013). http://ec.europa.eu/eurostat/en/web/products-statistical-books/-/KS-HA-13-001-04
- Feinberg, A. B., & Shapiro, E. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research*, 102(6), 453–462. https://doi.org/10.3200/JOER.102.6.453-462
- 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
- Holopainen, L., Kiuru, N., Mäkihonko, M., & Lerkkanen, M.-K. (2017). The role of part-time special education supporting students with reading and spelling difficulties from grade 1 to grade 2 in Finland. *European Journal of Special Needs Education*, 33, 316–333. https://doi.org/10.1080/08856257.2017.1312798
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multi-dimensional view. *Reading & Writing Quarterly*, 25(1), 4–32. https://doi.org/10.1080/10573560802491208
- Ise, E., Blomert, L., Bertrand, D., Faísca, L., Puolakanaho, A., Saine, N., Surányi, Z., Vaessen, A., Csépe, V., Lyytinen, H., Reis, A., Ziegler, J., & Schulte-Körne, G. (2011). Support systems for poor readers: Empirical data from six EU member states. *Journal of Learning Disabilities*, 44(3), 228–245. https://doi.org/10.1177/0022219410374235
- Itkonen, T., & Jahnukainen, M. (2010). Disability or learning difficulty? Politicians or educators? Constructing special education in Finland and the United States. *Comparative Sociology*, 9(2), 182–201. https://doi.org/10.1163/156913210X12536181351033
- Johnson, E. S., Jenkins, J. R., & Petscher, Y. (2010). Improving the accuracy of a direct route screening process. Assessment for Effective Intervention, 35(3), 131–140. https://doi.org/10.1177/1534508409348375
- Karing, C., & Artelt, C. (2013). Genauigkeit von Lehrpersonenurteilen und Ansatzpunkte ihrer förderung in der Ausund Weiterbildung von Lehrkräften. *Beiträge zur Lehrerbildung*, 31(2), 166–173.
- Kent, S. C., Wanzek, J., & Yun, J. (2018). Screening in the upper elementary grades: Identifying fourth-grade students at-risk for failing the state reading assessment. *Assessment for Effective Intervention*, 44, 160–172. https://doi.org/10.1177/1534508418758371
- Kikas, E., Soodla, P., & Mägi, K. (2017). Teacher judgments of student reading skills: Associations with child- and classroom related factors. *Scandinavian Journal of Educational Research*, 62(5), 783–797. https://doi.org/10. 1080/00313831.2017.1307271
- Kim, Y.-S., Petscher, Y., & Foorman, B. (2015). The unique relation of silent reading fluency to end-of-year reading comprehension: Understanding individual differences at the student, classroom, school, and district levels. *Reading and Writing*, 28(1), 131–150. https://doi.org/10.1007/s11145-013-9455-2
- Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45(2), 230–251. https://doi.org/10.1598/ rrq.45.2.4
- Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J.-E. (2008). Letter knowledge predicts grade 4 reading fluency and reading comprehension. *Learning and Instruction*, 18(6), 548–564. https://doi.org/10.1016/j.learninstruc.2007.11. 004
- Lerkkanen, M.-K. (2007). The beginning phases of reading literacy instruction in Finland. In P. Linnakylä, & I. Arffman (Eds.), *Finnish reading literacy. When quality and equity meet* (pp. 155–174). University of Jyväskylä, Institute for Educational Research.
- Lerkkanen, M.-K., Ahonen, T., & Poikkeus, A.-M. (2011). The development of reading skills and motivation and identification of risk at school entry. In M. Veisson, E. Hujala, P. K. Smith, M. Waniganayake, & E. Kikas (Eds.), Global perspectives in early childhood education: Diversity, challenges and possibilities (pp. 237–238). Peter Lang.
- Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, M., Siekkinen, M., & Nurmi, J.-E. (2006). *The first steps study [Alkuportaat], ongoing.* University of Jyväskylä, Finland.
- Lindeman, J. (1998). ALLU- Ala-asteen lukutesti [ALLU Reading test for primary school]. University of Turku, Finland: The Center for Learning Research.
- Lyytinen, H., Erskine, J., Tolvanen, A., Torppa, M., Poikkeus, A.-M., & Lyytinen, P. (2006). Trajectories of reading development: A follow-up from birth to school age of children with and without risk for dyslexia. *Merrill-Palmer Quarterly*, 52(3), 514–546. https://doi.org/10.1353/mpq.2006.0031
- Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, 102(3), 701–711. https://doi.org/10.1037/a0019135
- Marchand, G. C., & Furrer, C. J. (2014). Formative, informative, and summative assessment: The relationship among curriculum-based measurement of reading, classroom engagement, and reading performance. *Psychology in the Schools*, 51(7), 659–676. https://doi.org/10.1002/pits.21779



- Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools*, 48(4), 343–356. https://doi.org/10.1002/pits.20558
- Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, 14(2), 78–102. https://doi.org/10.1080/10627190903039429
- Mayringer, H., & Wimmer, H. (2003). SLS 1-4: Das Salzburger Lese-Screening für die Klassenstufen 1-4. Verlag Hans Huber.
- Meisinger, E., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of Dyslexia*, 60(1), 1–17. https://doi.org/10.1007/s11881-009-0031-z
- Oslund, E. L., Clemens, N. H., Simmons, D. C., & Simmons, L. E. (2018). The direct and indirect effects of word reading and vocabulary on adolescents' reading comprehension: Comparing struggling and adequate comprehenders. *Reading and Writing*, 31(2), 355–379. https://doi.org/10.1007/s11145-017-9788-3
- Paleczek, L., Seifert, S., & Gasteiger-Klicpera, B. (2017). Influences on teachers' judgment accuracy of reading abilities on second and third grade students: A multilevel analysis. *Psychology in the Schools*, 54(3), 228–245. https://doi.org/10.1002/pits.21993
- Paleczek, L., Seifert, S., Schwab, S., & Gasteiger-Klicpera, B. (2015). Assessing reading and spelling abilities from three different angles—correlations between test scores, teachers' assessment and children's self-assessments in L1 and L2 children. *Procedia-Social and Behavioral Sciences*, 174, 2200–2210. https://doi.org/10.1016/j.sbspro. 2015.01.876
- Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading*, 16(3), 187–217. https://doi.org/10.1080/10888438.2010.538780
- Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47(6), 427–469. https://doi.org/10.1016/j.jsp.2009.07.001
- Solis, M., Miciak, J., Vaughn, S., & Fletcher, J. M. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disability Quarterly*, *37*(4), 218–229. https://doi.org/10.1177/0731948714528806
- Soodla, P., & Kikas, E. (2010). Teachers' judgment of students' reading difficulties and factors related to its accuracy. In A. Toomela (Ed.), *Systemic person-oriented study of child development in early primary school* (pp. 73–94). Peter Lang.
- Soodla, P., Tammik, V., & Kikas, E. (2019). Is part-time special education beneficial for children at risk for reading difficulties? An example from Estonia. *Dyslexia*, 1–25. https://doi.org/10.1002/dys.1643
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42(8), 795–819. https://doi.org/10.1002/pits.20113
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. https://doi.org/10.1037/a0027627
- Takala, M., Wikman, K., Uusitalo-Malmivaara, L., & Lundström, A. (2015). Becoming a special educator: Finnish and Swedish students' views of their future professions. *Educational Inquiry*, 6(1), 25–51. https://doi.org/10.3402/edui. v6.24329
- 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
- Virinkoski, R., Lerkkanen, M.-K., Holopainen, L., Eklund, K., & Aro, M. (2018). Teachers' ability to identify children at early risk for reading difficulties in grade 1. *Early Childhood Education Journal*, 46(5), 497–509. https://doi.org/10.1007/s10643-017-0883-5
- Wagner, D. L., & Espin, C. A. (2015). The reading fluency and comprehension of fifth- and sixth-grade struggling readers across brief tests of various intervention approaches. *Reading Psychology*, 36(7), 545–578. https://doi.org/10.1080/02702711.2014.927812
- Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, 23(8), 889–912. https://doi.org/10.1007/s11145-009-9179-5
- Wayman, M. M., Wallace, T., Wiley, H. I., Tichdt, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, 41(2), 85–120. https://doi.org/10.1177/00224669070410020401
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson® III test. Riverside Publishing Company. Woolley, G. (2008). The assessment of reading comprehension difficulties for reading intervention. Australian Journal of Learning Difficulties, 13(1), 51–62. https://doi.org/10.1080/19404150802093729
- Zumeta, R. O., Compton, D. L., & Fuchs, L. S. (2012). Using word identification fluency to monitor first-grade reading development. *Exceptional Children*, 78(2), 210–220. https://doi.org/10.1177/001440291207800204



III

DEVELOPMENT OF READING AND ARITHMETIC SKILLS ACROSS GRADES 1 TO 4 IN TWO GROUPS OF CHILDREN RECEIVING PART-TIME SPECIAL EDUCATION

by

Riitta Virinkoski, Kenneth Eklund, Marja-Kristiina Lerkkanen, Leena Holopainen & Mikko Aro, 2020

Learning and Individual Differences

https://doi.org/10.1016/j.lindif.2020.101956

Reproduced with the kind permission of Elsevier.



Contents lists available at ScienceDirect

Learning and Individual Differences

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



Development of reading and arithmetic skills across Grades 1 to 4 in two groups of children receiving part-time special education[☆]

Riitta Virinkoski^{a,*}, Kenneth Eklund^b, Marja-Kristiina Lerkkanen^c, Leena Holopainen^d, Mikko Aro

- ^a Open University, Department of Teacher Education, P.O. Box 35, 40014, University of Jyväskylä, Finland
- ^b Faculty of Education and Psychology, P.O. Box 35, 40014, University of Jyväskylä, Finland
- ^c Department of Teacher Education, P.O. Box 35, 40014, University of Jyväskylä, Finland
- ^d School of Educational Sciences and Psychology, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland
- ^e Department of Education, P.O. Box 35, 40014, University of Jyväskylä, Finland

ARTICLE INFO

Keywords: Part-time special education Reading fluency Reading comprehension Arithmetic fluency Overlapping difficulties

ABSTRACT

This study investigated why some Finnish students receive part-time special education in Grade 1, duration of that support, and its relation to student reading fluency, reading comprehension, and arithmetic fluency development. The participants included two groups from Grade 1 receiving part-time special education (1-2 years vs. 3-4 years) and a control group comprising the remaining participants from the study. Teachers identified reading and expressive language difficulties as the main reasons for part-time special education in Grade 1. By Grade 4, students who received support until Grade 1 or 2 caught up the level of the control group especially in reading fluency. Students who had received support until Grade 3 or 4 exhibited more persistent and overlapping difficulties with attention, receptive language, memory, and motivation. Additionally, by Grade 4, differences from the control group remained statistically significant for all three skills. Implications for enhancing special education interventions for students with persistent and overlapping learning difficulties are discussed.

1. Introduction

Children face many learning challenges on commencing school, and difficulties in one academic domain have been found to increase the risk of difficulties in another (Cirino et al., 2018; Koponen et al., 2018; Landerl & Moll, 2010). For instance, comorbid difficulties in reading and mathematics are more common than might be expected statistically from the prevalence of deficits in single domains (e.g., Koponen et al., 2018; Landerl & Moll, 2010). To prevent more severe or prolonged learning difficulties, students should receive appropriate support as soon as any such difficulties are identified. In the Finnish educational context, primary school classroom teachers are the first to provide general support for learning reading and mathematics. If classroom teachers' support proves insufficient for students with reading, spelling, and arithmetic difficulties, part-time special education is provided individually or in small groups by special educational needs (SEN) teachers or special educators as a pull-out service, usually in their resource room (Holopainen et al., 2017; Lemons et al., 2018; Soodla et al., 2019; Takala & Ahl, 2014). In international comparisons, part-time special education for students with learning difficulties is quite extensive in Finland; in 2017, 22% of all students in Grades 1 to 9 participated in part-time special education at some point in the school year (Official Statistics of Finland, 2018). However, municipalities differ in terms of the share of students who receive special education and how such support is organized (Pulkkinen et al., 2019). The present study explores the reasons for providing part-time special education at the beginning of the school path, as well as the duration of the support period and how it affects subsequent skills development. To that end, the study compared two groups of students receiving part-time special education for differing lengths of time and followed the development of their academic skills until Grade 4.

1.1. Developing academic skills and overlapping difficulties

In Finland, pre-literacy skills are supported during compulsory preprimary education for 6-year-old children. Although pre-primary

^{*} This study was financed by the Academy of Finland (No. 292466 for 2015–2019; No. 268586 for 2013–2017).

Corresponding author at: Open University, P.O. Box 35, 40014, University of Jyväskylä, Finland. E-mail address: riitta.e.virinkoski@jyu.fi (R. Virinkoski).

education does not include formal instruction in literacy or math, children are introduced to letter names (Lerkkanen et al., 2004), mathematical concepts, and numbers (Aunola et al., 2004). Formal instruction in reading, spelling, and arithmetic starts at age seven, when children enter the first grade. The orthographic regularity of the Finnish language enables word reading and spelling to be taught simultaneously to first graders in the knowledge that one will complement the development of the other (Lerkkanen, 2007). As one of the most transparent alphabetic orthographies (Aro, 2017), Finnish supports rapid mastery of decoding accuracy. About 30% of first grade students can decode before school entry, and the rest acquire good decoding skills within months of beginning reading instruction at school (Lerkkanen et al., 2004). For that reason, reading difficulties (RD) in Finnish relate mainly to problems of reading fluency after learning to decode (Torppa et al., 2015), and these difficulties have been shown to persist (Eklund et al., 2015).

Fluent readers are able to read texts accurately at a conversational rate with appropriate prosody (Torgesen & Hudson, 2006). To comprehend texts effectively, one must also be able to understand the meanings of the words (Hudson et al., 2008; Mancilla-Martinez & Lesaux, 2010). In addition to difficulties with word recognition, reading comprehension issues relate to the nature of the text, insufficient vocabulary skills, and the objectives of reading (Mancilla-Martinez & Lesaux, 2010; Van der Lely & Marshall, 2010). For some students, poor word reading and comprehension skills can present obstacles to learning throughout compulsory education and into adulthood (Holopainen & Hakkarainen, 2019; Landerl & Wimmer, 2008). In a recent German follow-up study (N = 4123) across Grades 3 to 5, Moll et al. (2020) noted the prevalence of stability of fluency difficulties, which are hard to remediate.

The development of basic mathematical skills is a hierarchical process comprising basic number knowledge, verbal counting skills, memorizing arithmetic facts, understanding and using basic arithmetic operations (addition, subtraction, multiplication and division), understanding concepts, and being able to follow relevant procedures (Butterworth, 2005; Geary et al., 2012; Jordan et al., 2009; Zhang et al., 2020). Rapid retrieval of arithmetic facts and the ability to calculate by deploying such facts is the basis of math fluency. When a student must consistently rely on counting-based strategies (e.g., solving 4 + 3 by counting "four, five, six, seven"), arithmetic skills remain dysfluent (Koponen et al., 2018). Morgan et al. (2009; see also Nelson & Powell, 2018) reported that the stability of math difficulty from kindergarten to Grade 5 was 28%–65%, where the higher rate related principally to the lowest achieving students (≤10th percentile).

As noted above, learning difficulties exhibit strong comorbidity; for example, students with RD also tend to have difficulties with spelling (Holopainen et al., 2017; Landerl & Moll, 2010; Landerl & Wimmer, 2008), mathematics (Child et al., 2019; Cirino et al., 2015; Koponen et al., 2018; Moll et al., 2019), or language (Baird, 2008; Isoaho et al., 2016; Spanoudis et al., 2019) more frequently than expected. Many studies have also identified comorbidity of reading and attention difficulties (Grills-Taquechel et al., 2013; Peterson et al., 2017; Prochnow et al., 2013) and co-occurrence of reading and math difficulties with attention difficulties (Child et al., 2019). There is also some evidence that overlapping difficulties are especially persistent; for example, (Koponen et al., 2018) showed that comorbid difficulties in reading and arithmetic are relatively stable after Grade 2 and are more persistent than single-domain difficulties throughout the individual's school years. Korpipää et al., (2017) reported time-general intercorrelations between reading and arithmetic skills in Grades 1 and 7, and there is evidence that overlapping difficulties of this kind can impair students' academic careers beyond Grade 9 (Holopainen & Hakkarainen, 2019).

1.2. Support for learning difficulties and its effect on students' skills development

It has been reported that classroom teachers tend to provide more

instructional support and more active instruction to lower-performing students (Griffiths & Stuart, 2013; Kikas et al., 2015). However, some students with early learning difficulties subsequently require more support than classroom teachers can provide. As Finland's comprehensive part-time special education system is needs-based and free of charge, "low-threshold" learning support can be provided to any student as soon as there is an observed need. For students with milder learning or behavioral difficulties, no formal diagnosis is required (Björn et al., 2016; Jahnukainen & Itkonen, 2016). In some cases, a student with persistent wide-ranging difficulties in reading, spelling, and arithmetic (or more severe learning difficulties in one skill) will receive regular or irregular support from an SEN teacher during their early school years (see Holopainen et al., 2017).

Expressive or receptive language disorders (see Snowling & Melby-Lervåg, 2016; Spanoudis et al., 2019) and difficulties in reading, spelling, and arithmetic are the most typical reasons for providing special education support at the beginning of school (Holopainen et al., 2017; Lappalainen & Hotulainen, 2012; Morgan et al., 2011). To prevent severe difficulties, students should receive support as early as possible; in Finland, the decision to provide support is not based on national assessments, which do not exist. Instead, teacher evaluations determine which students are in need of support ad which support practices are selected. These decisions are based on multi-professional evaluation and take account of the views of SEN teachers, regular teachers, and parents (Björn et al., 2016; Takala et al., 2009). Support may be implemented through co-teaching in regular classrooms (push-in)-for example, by the classroom teacher and SEN teacher-or in a separate class, individually or in small groups (3-4 students) with the SEN teacher (pullout), for as long as the student needs it (Holopainen et al., 2017; Takala et al., 2009).

Part-time special education is most extensively provided at the beginning of school as a preventive measure to ensure that students learn to read and spell (Takala & Ahl, 2014; Takala et al., 2009). In the case of reading, support is usually provided in the form of extra lessons (usually 1–2 h per week), including training in reading, spelling, and phonological awareness (Holopainen et al., 2017; Ise et al., 2011). While some students seem to benefit from a short period of support, others may have deeper and wider learning difficulties that cannot be remediated by short-term or temporary support (Holopainen et al., 2017). Previous studies have shown that pull-out instruction in a resource-type room often proves beneficial for students with learning difficulties (Bottge et al., 2018; Lemons et al., 2018), especially if the group is small (five or fewer students) (Vaughn & Wanzek, 2014).

However, regardless of the intensity or quality of the intervention, it is widely reported that some students fail to make adequate progress, and they lag behind the expected grade-level performance (Holopainen et al., 2017; Vaughn & Wanzek, 2014). Several studies have noted the issue of low responsiveness or non-responsiveness to interventions, highlighting the need for more individualized interventions and practices and for more precise definition of the duration of support (e.g., Denton, 2012; Denton et al., 2006; Miciak et al., 2017; Vaughn & Wanzek, 2014). With regard to overlapping difficulties, there is also evidence that students with co-occurring learning difficulties and attention deficit hyperactivity disorder (ADHD) exhibit a lack of responsiveness to instruction or support (DuPaul et al., 2013; Vaughn & Fuchs, 2003).

1.3. Study aims and context

The aim of the present study was to investigate SEN teachers' evaluations of Grade 1 students' special education needs, and how duration of support related to progress in reading fluency, reading comprehension, and arithmetic fluency in Grades 1 to 4. One of the two participating groups had received early-stage support (ESS) for 1–2 years while the other had received long-lasting support (LLS) for 3–4 years. A large population sample was used as the control group. The study sought to

identify any differences between the two support groups in terms of reported reasons for part-time special education and additional difficulties in Grade 1. We also compared all three groups' reading and arithmetic skills at the beginning and end of the follow-up period (in Grades 1 and 4) and examined how these skills developed.

The study augments the earlier findings of Holopainen et al. (2017) and Soodla et al. (2019) regarding the role of part-time special education support for students with reading difficulties (RD) across Grades 1 and 2. In a follow-up study, Holopainen et al. (2017) compared the development of reading and spelling skills of students with RD, who received part-time special education, and students without RD, who did not. In Grade 1, students with RD lagged behind their age level, but the gap between the groups had decreased by Grade 2. Comparison of students with RD alone and students with overlapping difficulties showed that the latter group was slower to develop. The preceding study also investigated the frequency and form (individual, small group, or both) of part-time special education provision and how these related to the development of reading and spelling skills in students with RD alone and students with RD and overlapping difficulties. The results indicated that frequent individual provision was associated with slower skill development while skills developed faster with moderate frequency of provision in small groups (3-4 students). In another follow-up study, Soodla et al. (2019) found that one year of part-time special education (two lessons per week) may not be the most effective form of support for students with RD, especially when that support is provided without prior diagnostic assessment, targeted intervention, or frequent progress monitoring.

Based on previous research (Lappalainen & Hotulainen, 2012; Morgan et al., 2011), we anticipated that difficulties in reading, arithmetic, and expressive language were likely to be the most common reasons for support, and that the reasons would be quite similar in the two groups. On that basis, we formulated the following research question.

RQ1. What reasons did SEN teachers report for providing part-time special education in Grade 1, and to what extent did the reasons for support differ between the ESS and LLS groups?

We further expected that RD would overlap with arithmetic difficulty (Cirino et al., 2015; Moll et al., 2019), expressive language difficulty (Isoaho et al., 2016; Snowling & Melby-Lervåg, 2016), and attention difficulty (Grills-Taquechel et al., 2013; Prochnow et al., 2013), and that students in the LLS group would have more additional difficulties. On that basis, we formulated the following research question.

RQ1a. What kinds of additional learning-related difficulties did SEN teachers report in Grade 1, and to what extent did the ESS and LLS groups differ in terms of the extent and characteristics of those difficulties?

We also anticipated that the ESS and the LLS groups would perform less well in reading than the control group, and that the ESS group would outperform the LLS group (Holopainen et al., 2017). On that basis, we formulated the following research question.

RQ2. Did the two support groups (ESS and LLS) differ in reading fluency, reading comprehension, and arithmetic fluency at the beginning of part-time special education in Grade 1 compared to their peers in the large follow-up sample?

Again based on prior research (Holopainen et al., 2017), we anticipated that skills development among the LLS group would be slower than among the ESS and control groups, and that the ESS group would outperform the LLS group on every skill at the end of the follow-up period. We also expected that the ESS group would catch up with the control group. On that basis, we formulated the following research question.

RQ3. Did the development of reading fluency, reading comprehension, and arithmetic fluency skills in the ESS and LLS groups across Grades 1 to 4 (as compared to their peers in the large follow-up sample) reflect the duration of part-time special education, and to what extent did the ESS group, the LLS group, and the control group differ in terms of those skills at the end of the follow-up?

2. Method

2.1. Participants and procedure

The study data were drawn from an extensive follow-up study (Lerkkanen et al., 2006-2016) in four municipalities in different parts of Finland, involving five data collection time points: fall 2007 and spring 2008 (Grade 1), spring 2009 (Grade 2), spring 2010 (Grade 3), and spring 2011 (Grade 4). The participants included one whole age cohort of children in three of the participating municipalities and half of the age cohort in the fourth municipality. Parental education levels were similar to the Finnish national average (Eurostat Regional Yearbook, 2013). Informed consent for participation was collected from the students' teachers and caretakers. Within the sample (N = 2157), risk for RD was determined at the end of the kindergarten year on the basis of four criteria: initial phoneme identification (indicating phonological awareness), letter knowledge, rapid automatized naming, and self-reported parental RD (see Lerkkanen et al., 2011). RD risk was defined as joint occurrence of at least two of three criteria: low phonological awareness (clearly below age level score for initial phoneme identification, <15th percentile); poor letter knowledge (≤15th percentile); and poor rapid automatized naming (≤15th percentile) (Lerkkanen et al., 2011). If parents themselves reported having reading difficulties, a score below the 15th percentile in any one of the three tests (phonological awareness, letter knowledge, or rapid automatized naming) was considered a sufficient indicator of RD risk. Control group members were randomly selected from the same classrooms as those identified as at risk for RD.

Of the 2157 children in the sample, a sub-study followed 598 (53% boys, 47% girls; 277 at risk for RD, 321 not at risk for RD) individually from Grade 1. In Grade 1, 35 SEN teachers (94% female; mean age = 45.6 years; SD = 9.60) from four municipalities participated in the special education sub-study; issues investigated included pedagogical practices for children with learning difficulties. The SEN teachers received a list of children who were followed more intensively, but they did not know which children were at risk for RD and which were not. The teachers were asked to rate those who had received special education in Grade 1; there were typically two or three such students, but if the number exceeded six, the teachers were allowed to select students to be rated. Each year, the SEN teachers were asked to complete individual student reports and questionnaires regarding special education content and implementation. For present purposes, we used the questionnaires and reports collected during Grade 1 (fall and spring terms).

The support groups comprised 130 children who had received parttime special education during the fall and/or spring term in Grade 1. We assigned the students to two groups: the ESS group (n=55, 64% boys), who had received support in Grade 1, Grade 2, or both, and the LLS group (n=75, 69% boys), who had received support from Grade 1 until Grade 3 or 4. The control group (n=2027, 52% boys) comprised peers who participated in the follow-up but were not included in the two support samples. Control group performance was used as an indicator of age-appropriate skills development. The research follows the principles that are endorsed by the research community, that is, integrity, meticulousness, and accuracy in conducting research, and in recording, presenting, and evaluating the research results. In the treatment of our samples research ethical standards and in reporting publishing ethics have been complied. The study has been reviewed and approved by the Ethical Board of the University of Jyväskylä in 2006.

2.2. Measures

2.2.1. Teacher ratings of students' learning difficulties

Learning difficulties were identified on the basis of SEN teachers' reports in Grade 1. The questionnaires (sent in the fall and spring terms) asked teachers to specify the reasons for part-time special education provision by responding to the following open-ended question: "What is the main reason for support?" Responses were assigned to four main categories (Reading difficulty, Expressive language difficulty, Arithmetic difficulty, Other difficulty) and 11 sub-categories (see Table 1). Sometimes the teachers reported a student having several reasons for providing support, and categories were formed accordingly. In addition to the main reason(s) for receiving part-time special education the teachers were asked to identify each individual student's other learning related difficulties as follows: "Does the student receiving part-time special education have other difficulties?" Based on a list of 10 options (1 = Receptive language difficulty, 2 = Expressive language difficulty, 3 = Perception difficulty, 4 = Memory difficulty, 5 = Attention difficulty, 6 = Motor difficulty, 7 = Motivation difficulty, 8 = Socio-emotional difficulty, 9 = Behavior difficulty, and 10 = Other: what?), the teachers were asked to rate each student on a three-point scale (0 = not at all, 1 = mild problems/ sometimes, 2 = clear problems/often). In the analysis, mild problems (or sometimes) and clear problems (or often) were pooled, and scores were summed to determine the total number of observed difficulties.

Students' reading fluency, reading comprehension, and arithmetic fluency in Grades 1 to 4 were assessed annually during the spring term using identical tests, which were group-administered in the classroom by trained research assistants.

2.2.2. Reading fluency

Two tasks were used to test reading fluency: ALLU TL2 (Lindeman, 1998) and Test of Silent Reading Efficiency and Comprehension (TOS-REC; Wagner et al., 2009; Finnish adaptation by (Lerkkanen & Poikkeus, 2009). A group-administered word reading fluency task, ALLU is a subtest of a standardized Finnish reading test battery for primary schools (ALLU; Lindeman, 1998). Each of the 80 items comprises one picture and four phonologically similar words that are familiar and frequently

Table 1Main reasons for part-time special education provision for Grade 1 students (n = 130) as reported by SEN teachers.

Main reason for support	Student sample $(n = 130)$ %	ESS group $N = 55\%$ (n)	LLS group N = 75% (n)
Reading difficulty (RD): total	69		
RD alone		44 (24)	39 (29)
RD with expressive language		5 (3)	7 (5)
difficulty			
RD with arithmetic difficulty		2(1)	7 (5)
RD with attention difficulty		4(2)	7 (5)
RD with some other difficulty ^a		7 (4)	16 (12)
Expressive language difficulty ^a : total	30		
Expressive language difficulty alone		31 (17)	17 (13)
Expressive language difficulty with		5 (3)	7 (5)
RD			
Expressive language difficulty with		2(1)	0 (0)
attention difficulty			
Arithmetic difficulty: total	6		
Arithmetic difficulty alone		0 (0)	1(1)
Arithmetic difficulty with RD		2(1)	7 (5)
Arithmetic difficulty with some other		2(1)	0 (0)
difficulty ^b			
Other difficulty ^a , total	5	4 (2)	7 (5)

The sum exceeds 100% because teachers reported that most students had several difficulties.

used by children. Students were instructed to read the four words silently before matching the correct word to the picture by drawing a line. Scores were calculated as the number of correct responses within a two-minute time limit (maximum = 80). TOSREC (Wagner et al., 2009) is a sentence verification task comprising 60 sentences. The students were instructed to decide whether the content of each silently read sentence was true or false by circling the appropriate option (true/false). Scores were calculated as the number of correct responses within a three-minute time limit (maximum = 60). A composite score for reading fluency was derived separately for each grade by calculating the mean of the two standardized scores (ALLU TL2 and TOSREC). In Grades 1, 2, 3, and 4, Cronbach's alpha values were 0.86, 0.82, 0.82, and 0.83, respectively.

2.2.3. Reading comprehension

A standardized subtest of ALLU (Lindeman, 1998) was used to assess reading comprehension, with texts and questions customized for each grade level. The students were instructed to read a short story silently before either answering 12 multiple-choice questions (4 alternative responses) about the text or arranging statements by their order of appearance in the text. One point was awarded for each correct response (maximum = 12); students were allowed to complete the task at their own pace, up to a maximum of 45 min. Cronbach's alpha values were 0.98 in Grade 1, 0.94 in Grade 2, 0.96 in Grade 3, and 0.85 in Grade 4. Task scores were standardized separately for the four grade levels to enable assessment of progress in the ESS and LLS groups as compared to the control group.

2.2.4. Arithmetic fluency

Arithmetic fluency was tested using the Basic Arithmetic Test in Grades 1, 2, and 3 and the Basic Arithmetic II Test (Aunola & Räsänen, 2007) in Grade 4; the latter included more demanding items than in Grades 1 to 3. Both versions comprised 28 items, including addition and subtraction items in Grades 1, 2 and 3, and addition, subtraction, division, and multiplication items in Grade 4. Scores were calculated as the number of correct responses within a three-minute time limit (maximum = 28). Cronbach's alpha values in Grades 1, 2, 3, and 4 were 0.98, 0.86, 0.68, and 0.85, respectively. Scores were standardized within grade levels to enable assessment of progress in the ESS and LLS groups as compared to the control group.

2.2.5. Data analyses

IBM's SPSS Statistics 26 program was used to perform the statistical analyses. Descriptive statistics (frequencies, percentages, means, and standard deviations) were used to specify learning difficulties in Grade 1, using cross-tabulation with χ^2 tests to identify any differences between ESS and LLS group skills in Grade 1. Independent samples ttesting was used to assess any differences between the two groups in terms of additional difficulties in Grade 1. Analysis of variance was performed in Grade 4 to identify any differences between the two support groups and the control group in terms of initial skills (reading fluency, reading comprehension, and arithmetic fluency) in Grade 1, as well as final skill levels at the last time point. Using Cohen's d (Cohen, 1992), effect sizes were calculated to estimate the magnitude of any differences between the three groups (0.20 = small, 0.50 = medium, 0.80 = large). Mixed-design ANOVAs were used to investigate differences in development of these skills among the three groups. Grade level (1, 2, 3, 4) was used as the within-subjects factor, and group (ESS, LLS, control) served as the between-subjects factor. In all analyses of academic skills, standardized scores were used to demonstrate relative differences between the two support groups (ESS and LLS) as compared to the control group. The use of standardized scores also enabled group comparisons of skills (reading fluency, reading comprehension, and arithmetic fluency). As standardization was performed separately for each grade and measure, the main effect of grade is not reported in the mixed-design ANOVAs.

 $^{^{\}rm a}$ Expressive language difficulty = e.g., distortions of /s/ or /r/ sounds.

^b Other difficulty = behavior difficulty, motor difficulty, memory difficulty, socioemotional difficulty, perception difficulty, slowness in learning.

3. Results

3.1. Reasons for part-time special education provision and students' additional difficulties

To begin, we investigated SEN teachers' reported reasons for part-time special education provision in Grade 1, and the extent to which those reasons differed for the ESS and LLS groups. Reasons were assigned to four main categories and 11 sub-categories (see Table 1). The teachers' reports showed, first, that there were various reasons for part-time special education provision in Grade 1, and that some students required support for several reasons. The two most frequently mentioned reasons were reading difficulty (RD; 69%) and expressive language difficulty (30%); the least frequent was arithmetic difficulty (6%). For both ESS and LLS groups, the most frequent reason for part-time special education was RD, followed by expressive language difficulty.

We then examined students' additional learning-related difficulties in Grade 1 as reported by SEN teachers, and any differences between ESS and LLS groups in the reported extent of those difficulties (see Table 2). The most frequently reported additional difficulty was attention difficulty (64% in the ESS group versus 85% in the LLS group). Memory difficulty was the second most common additional difficulty in both support groups (62% in the ESS group versus 79% in the LLS group).

We were also interested in how the ESS and LLS groups differed in terms of additional difficulties. While the vast majority of students in both groups had several additional difficulties, the percentage of students with each difficulty was systematically higher in the LLS group (with the exception of socio-emotional difficulty). The χ^2 test revealed that the two groups differed in terms of the frequency of teacherreported difficulties in four categories: attention difficulty (χ^2 (1) = 8.2, adjusted standardized residual = 2.9, p = .004), receptive language difficulty (χ^2 (1) = 5.4, adjusted standardized residual = 2.3, p = .02), memory difficulty (χ^2 (1) = 4.4, adjusted standardized residual = 2.1, p = .04), and motivation difficulty (χ^2 (1) = 4.2, adjusted standardized residual = 2.1, p = .04). Based on the teachers' reports, students in the ESS group had an average of 4.67 difficulties (SD = 2.69) while the LLS group averaged 5.55 difficulties (SD = 2.39). Independent samples ttesting indicated that the difference between the two groups in terms of total difficulties was barely non-significant (t(128) = -1.954, p = .053).

Table 2Students' additional learning-related difficulties in Grade 1 as reported by SEN

Learning difficulty	ESS group $N = 55\%$ (n)	LLS group $N = 75\%$ (n)
Attention difficulty ^a	64 (35)	85 (64)
Memory difficulty ^b	62 (34)	79 (59)
Receptive language difficulty ^c	58 (32)	77 (58)
Motivation difficulty ^d	55 (30)	72 (54)
Perception difficulty ^e	53 (29)	65 (49)
Expressive language difficulty ^f	51 (28)	51 (38)
Motor difficulty ^g	44 (24)	55 (41)
Socio-emotional difficulty ^h	53 (29)	44 (33)
Behavior difficulty ⁱ	29 (16)	27 (20)

- ^a Concentration, impulsivity, executive functioning.
- ^b Auditory memory, forgetfulness.
- ^c Naming, vocabulary, comprehension of speech.
- d Work orientation, self-esteem, persistence.
- e Directions, left-right.
- f Articulation, stuttering, voice difficulty.
- g Fine motor skills, gross motor skills.
- h Timidity, anxiety, exhaustion, friendships.
- ⁱ Aggressiveness, defiance, harassment.

3.2. Differences in skill levels in Grade 1

We then investigated the extent to which the ESS and LLS groups differed in reading fluency, reading comprehension, and arithmetic fluency skills in Grade 1 when compared to the control group. The ANOVA results revealed significant differences between the groups in all tested skills (see Table 3). Post hoc tests with Bonferroni corrections showed that the LLS group differed from the control group in reading fluency (p < .001), reading comprehension (p < .001), and arithmetic fluency (p < .01). The difference between the ESS group and the control group was statistically significant for reading fluency (p = .001) and reading comprehension (p = .05) but not arithmetic fluency (p = .175). Finally, differences between the LLS and ESS groups were not significant for reading fluency (p = .114), reading comprehension (p = .161), or arithmetic fluency (p = 1.000).

Table 4 shows pairwise comparisons of group means and effect sizes (see Cohen, 1992) in Grade 1. The effect sizes of group differences between LLS and the control group were large for reading fluency and reading comprehension and small for arithmetic fluency. The effect sizes of differences between ESS and the control group were small for all three skills, as were the differences between the LLS and ESS groups.

3.3. Differences in skill development (Grades 1 to 4)

The third research question investigated differences in the development of reading fluency, reading comprehension, and arithmetic fluency between the three groups in Grades 1 to 4. Using mixed-design ANOVA, separate analyses were performed for each academic skill. Grade (1, 2, 3, and 4) was used as the within-subjects factor, and group (ESS group, LLS group, and control group) served as the between-subjects factor.

The mixed-design ANOVA for reading fluency identified a significant Grade x Group interaction effect $(F(6,\ 3710)=3.75,\ p<.01,\ \eta_P^2=0.006)$. To further assess differences between the groups in reading fluency development between Grades 1 and 2, Grades 2 and 3, and Grades 3 and 4, we tested for within-subject contrasts in Grade x Group interaction. The effect was significant for development between Grades 2 and 3 $(F(2,1857)=3.70,\ p<.05,\ \eta_P^2=0.004)$ but not for any other period, indicating differences in reading fluency development between Grades 2 and 3. Post hoc pairwise comparisons (see Fig. 1) of reading fluency improvement between Grades 2 and 3 showed that the ESS group had made greater gains in reading fluency than the control group $(F(1,1873)=6.89,\ p<.01,\ \eta_P^2=0.004)$ or the LLS group $(F(1,125)=5.57,\ p<.05,\ \eta_P^2=0.04)$, but there was no difference in reading fluency development between the LLS group and the control group $(F(1,1896)=0.91,\ p=.76,\ \eta_P^2=0.000)$.

In the mixed-design ANOVA for reading comprehension, the main effect of Group was significant ($F(2,1817)=28.16,p<.001,\eta_P^2=0.03$). The Grade x Group interaction effect was non-significant ($F(6,3630)=1.96,p=.07,\eta_P^2=0.003$), indicating similar development in all groups. In pairwise post-hoc analyses with Bonferroni correction (see Fig. 2), the control group exhibited better reading comprehension skills than the LLS (p<.001) and ESS (p<.05) groups. In addition, the ESS group outperformed the LLS group in reading comprehension (p<.05).

Finally, in the mixed-design ANOVA for arithmetic fluency, the main effect of Group was significant (F(2, 1848) = 15.56, p < .001, $\eta_P^2 = 0.017$), and there was a significant Grade x Group interaction effect (F(6, 3692) = 2.27, p < .05, $\eta_P^2 = 0.004$. To further assess differences between the groups in the development of arithmetic fluency across grade levels, we analyzed within-subjects contrasts in Grade x Group interaction and found a significant group difference in development between Grades 3 and 4 (F(2, 1848) = 3.27, p < .05), $\eta_P^2 = 0.004$). Post hoc pairwise comparisons with Bonferroni corrections (see Fig. 3) showed that arithmetic fluency had improved more in the LLS group than in the control group (F(1,1885) = 5.69, p < .05, $\eta_P^2 = 0.003$). There were no significant differences between the ESS group and the control group (F(1,1859) = 0.54, p = .46, $\eta_P^2 = 0.000$) or between the ESS and LLS groups

Table 3
Standardized scores and significant group differences in reading fluency, reading comprehension, and arithmetic fluency (Grade 1).

Skill	ESS group $n = 52-53$		LLS group $n = 74-75$		Control group $n = 1909-1924$		F(2, 2035–2049) ^a
	M	SD	M	SD	M	SD	
Reading fluency	-0.44 ^y	1.00	-0.81 ^y	0.78	0.04 ^z	0.99	32.28***
Reading comprehension Arithmetic fluency	$-0.37^{y} \\ -0.24^{y, z}$	0.99 1.01	$-0.71^{y} \\ -0.36^{y}$	0.84 0.91	0.04^{z} 0.02^{z}	0.99 1.00	24.09*** 6.78***

 $^{^{}m y,\ z}$ Groups with different superscripts differed significantly (p<.05).

Table 4Effect sizes of group differences in reading fluency, reading comprehension, and arithmetic fluency (Grade 1).

Skill	Control group vs. ESS group	Control group vs. LLS group	ESS group vs. LLS group	
	Effect size ^a	Effect size ^a	Effect size ^a	
Reading fluency	0.49	0.95	0.43	
Reading comprehension	0.41	0.81	0.39	
Arithmetic fluency	0.26	0.40	0.12	

^a Cohen's d (Cohen, 1992).

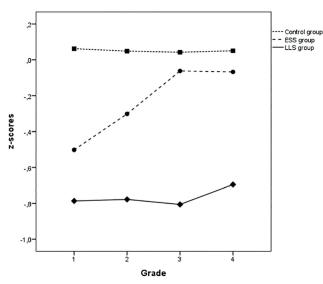
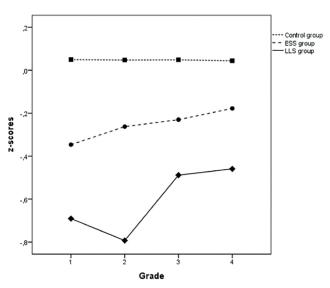


Fig. 1. Progress in reading fluency (Grades 1, 2, 3, and 4): control group, early-stage support (ESS) group, and long-lasting support (LLS) group.

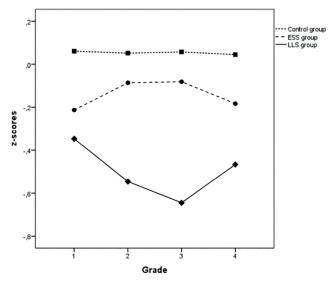
$$(F(1, 118) = 3.62, p = .06, \eta_P^2 = 0.03).$$

3.4. Skills differences in Grade 4

As well as investigating differences in the development of academic skills across the three groups, we also investigated differences in reading fluency, reading comprehension, and arithmetic fluency at the final assessment point at the end of Grade 4 (see Table 5). The ANOVA results indicated that the LLS group differed from the control group on all three skills. Post hoc tests with Bonferroni corrections showed that the LLS group differed from the control group in reading fluency (p < .001), reading comprehension (p < .001), and arithmetic fluency (p < .001). No differences were found between the ESS group and the control group. The differences between the ESS and the LLS groups were statistically significant for reading fluency (p < .01) but non-significant for reading



 $\begin{tabular}{ll} Fig.~2. Progress in reading comprehension (Grades~1,~2,~3,~and~4): control group, early-stage support (ESS) group, and long-lasting support (LLS) group. \\ \end{tabular}$



 $\begin{tabular}{ll} Fig.~3. Progress in arithmetic fluency (Grades~1,~2,~3,~and~4): control group, early-stage support (ESS) group, and long-lasting support (LLS) group. \\ \end{tabular}$

comprehension and arithmetic fluency.

Table 6 shows pairwise comparisons of group means and effect sizes (see Cohen, 1992) in Grade 4. The effect sizes of differences between the LLS group and the control group were large for reading fluency, medium

p < .001.

a Degrees of freedom vary between 2035 and 2049 due to missing values in dependent measures (reading fluency, reading comprehension, and arithmetic fluency).

Table 5
Standardized scores and significant group differences in reading fluency, reading comprehension, and arithmetic fluency in Grade 4.

Skill	ESS group LLS group $n = 47$ $n = 73$			Control group $n = 1829-1834$		F(2, 1949–1954) ^a	
	M	SD	M	SD	M	SD	
Reading fluency	$-0.09^{y,z}$	0.97	-0.69 ^y	0.96	0.03 ^z	0.99	19.04***
Reading comprehension	$-0.19^{y,z}$	1.05	-0.45^{y}	1.03	0.02^{z}	0.99	8.92***
Arithmetic fluency	$-0.17^{y,z}$	1.13	$-0.47^{y,z}$	0.97	0.02^{z}	0.99	9.22***

 $^{^{}m y,\ z}$ Groups with different superscripts differed significantly from each other (p < .05).

Table 6Effect sizes of group differences in reading fluency, reading comprehension, and arithmetic fluency in Grade 4.

Skill	Control group vs. ESS group	Control group vs. LLS group	ESS group vs. LLS group	
	Effect size ^a	Effect size ^a	Effect size ^a	
Reading fluency Reading comprehension	0.13 0.26	0.75 0.47	0.63 0.25	
Arithmetic fluency	0.19	0.50	0.29	

ESS = early-stage support group, LLS = long-lasting support group.

for arithmetic fluency, and small for reading comprehension. The effect sizes of differences between the ESS group and the control group were small for all three skills. The effect sizes of differences between the LLS and ESS groups were medium for reading fluency and small for reading comprehension and arithmetic fluency.

4. Discussion

4.1. General discussion

The study investigated the development of reading fluency, reading comprehension, and arithmetic fluency in two groups (ESS and LLS) that had received part-time special education support for 1–2 years and 3–4 years, respectively, as compared to a control group. The aim was to identify teacher-reported reasons for part-time special education provision, and how additional learning-related difficulties in Grade 1 pertained to duration of support. The study also examined the development across Grades 1 to 4 of reading fluency, reading comprehension, and arithmetic fluency and skill levels in the two support groups at two timepoints (Grade 1 and Grade 4).

The results indicate the differing developmental trajectories of the two support groups across Grades 1 to 4. By Grade 4, students in the ESS group had caught up with their control group peers, confirming the benefit of short support periods. These results align with earlier evidence that struggling students can usually improve their literacy skills with individual attention or small group intervention (Denton et al., 2006; Griffiths & Stuart, 2013; Kikas et al., 2015). While the need for support may decrease after Grade 1 or 2, our findings show that students with persistent and overlapping learning difficulties (the LLS group) generally received part-time special education support throughout their early school years. This suggests that some students may need more intensive and targeted intervention based on systematic assessment of core difficulties if they are to catch up with their typically achieving peers (see Denton, 2012; Denton et al., 2006). However, there is also evidence that frequent individual support of this kind (>48 h per year) may be associated with slower skills development (Holopainen et al., 2017).

4.2. Reported reasons for part-time special education; overlapping difficulties

As we anticipated, RD was the most frequently reported reason for provision of part-time special education for both support groups in Grade 1. This is understandable in light of the selection criteria for individual follow-up (i.e., identified risk for RD). According to Official Statistics of Finland (2018), RD is the most common reason for part-time special education at the beginning of primary school in Finland. As we hypothesized, expressive language difficulty alone was the second most common reported reason for support in both groups. Unexpectedly, however, arithmetic difficulty was relatively infrequently identified in both groups as a reason for part-time special education. It is conceivable that the tasks used to evaluate early-stage arithmetic skill acquisition may not reveal difficulties in fluency (Koponen et al., 2018), so complicating precise identification of the difficulty. Previous studies (e. g., Zhang et al., 2020) have also reported that difficulties in arithmetic fluency become apparent only after they begin to accumulate.

Although differences between the LLS and ESS groups in terms of additional learning-related difficulties reported by teachers did not quite reach the accepted significance level (0.05), the LLS group outnumbered the ESS group in four categories: attentional difficulty, receptive language difficulty, memory difficulty, and motivation difficulty. As we anticipated, the most frequently reported RD-related issue in both support groups was attention difficulty, and this aligns with earlier studies (Grills-Taquechel et al., 2013; Peterson et al., 2017; Prochnow et al., 2013). The inattentive subtype of ADHD (ADHD-I) is more strongly linked with difficulties in learning than hyperactivity-impulsivity (Massetti et al., 2008). In particular, children who exhibit inattentiveness at school entry continue to have poor literacy outcomes throughout elementary school (Prochnow et al., 2013) or poorer reading and mathematics outcomes when compared to a control group without ADHD (Massetti et al., 2008).

Contrary to expectation, there were no reports of overlap between arithmetic difficulty and the main reasons for part-time special education provision. This may reflect SEN teachers' lower preparedness to identify children struggling with math. In Finnish schools, learning to read and spell are generally considered the key skills to be acquired at the beginning of compulsory education (Takala & Ahl, 2014; Takala et al., 2009). Additionally, SEN teacher education typically emphasizes literacy assessment and interventions, which prepare teachers to support students with reading difficulties more than those struggling with arithmetic. The observed co-occurrence of RD and expressive or receptive language difficulty aligns with previous evidence (Cabbage et al., 2018; Isoaho et al., 2016; Skibbe et al., 2008; Snowling & Melby-Lervåg, 2016) regarding difficulties in developing academic skills. Although language impairment is known to be highly heterogeneous, it has been shown to constitute a significant risk of literacy learning difficulties for some school-age children (Isoaho et al., 2016).

There is also evidence that deficits in working memory contribute to poor school achievement, especially in reading and math, and co-occurrence with ADHD and working memory problems has been

ESS = early-stage support group, LLS = long-lasting support group.

^{***} p < .001

a Degrees of freedom vary between 1949 and 1954 due to missing values in dependent measures (reading fluency, reading comprehension, and arithmetic fluency).

^a Cohen's d (Cohen, 1992).

reported across Grades 2 to 4 (Maehler & Schuchardt, 2016). Low school motivation has also been linked to poor reading and math performance and low academic well-being, as well as to risk of eventual education dropout (Korhonen et al., 2014; Parhiala et al., 2018; Torppa et al., 2020), although these findings relate to older students (i.e., 15–16-year-olds).

Existing evidence that overlapping difficulties can obscure the main difficulty and so complicate targeted support highlights the importance of early identification, thorough evaluation of skills, and periodic individual assessment (Peterson et al., 2017). Students with co-occurring learning difficulties and ADHD have previously manifested a lack of responsiveness to support (DuPaul et al., 2013; Vaughn & Fuchs, 2003). Our study lends support to these findings and confirms that co-occurrence of learning and other difficulties seems to increase the risk of non-responsiveness to support. It has also been shown that students with overlapping difficulties experience poorer outcomes in all areas than peers whose difficulty is confined to a single domain (Cirino et al., 2015; Willcutt et al., 2013) and tend to respond inadequately to proven interventions (Fuchs et al., 2004).

4.3. Skills development: Grades 1 to 4

As expected, the ESS group differed somewhat from the control group in Grade 1 reading fluency and comprehension but not in arithmetic fluency. As also expected, the skills gap between ESS and control group narrowed, and there was no significant difference between them at the final follow-up point in Grade 4. One possible explanation for this positive development is that support could be accurately targeted because the most commonly reported reason for part-time special education provision in the ESS group was RD alone (44%) or expressive language difficulty alone (31%).

As hypothesized, the LLS group and the control group differed on every skill in Grade 1. Across Grades 1 to 4, the LLS group was slower than the control group in developing reading fluency, reading comprehension, and arithmetic fluency, and those differences persisted until Grade 4. Grade 1 effect sizes show that the two support groups differed in terms of starting points for reading fluency (0.43) and reading comprehension (0.39), both of which favored the ESS group. While there were no statistically significant differences between the ESS and LLS groups in Grade 1, the former had achieved better reading fluency by Grade 4. In consistent orthographies like Finnish, dysfluency is the most widely acknowledged indicator of RD risk (Eklund et al., 2015; Share, 2008; Solheim et al., 2020; Torppa et al., 2017). In such transparent orthographies, accurate decoding skill is relatively easy to master, even for those at risk for RD, but they may encounter persistent reading fluency difficulties in later grades (Torppa et al., 2017). Students in the ESS group exhibited reading difficulties on commencing school, but these were resolved once they learned to decode and achieved typical reading accuracy (see Torppa et al., 2015). In contrast, fluency difficulties are often persistent (Moll et al., 2020) and less easily remediated.

The slower development of academic skills in the LLS group can be partly explained by the greater number of reported reasons for part-time special education provision (e.g., RD with arithmetic difficulty, RD with some other difficulty) as compared to the ESS group. Although the ESS group exhibited significantly more difficulties related to attention, receptive language, memory, and motivation, there was no significant difference between the groups in terms of overlapping learning-related difficulties. To meet specific individual needs, the offered support must address a wide range of learning-related challenges, which explains the weaker response to support and the observed need to prolong that support. Recent studies have also shown that overlapping reading and arithmetic difficulties are more stable than single-domain difficulties (Cirino et al., 2018; Koponen et al., 2018), and Griffiths and Stuart (2013) and Soodla et al. (2019) have suggested that efficacy of intervention is not necessarily linked to duration. Other studies (e.g., Soodla et al., 2019; Vaughn & Wanzek, 2014; Wanzek et al., 2014) have also reported that the persistent nature of some students' learning challenges makes these difficult to remediate, even with special educational support over several years.

While all students who show early signs of learning difficulties can be provided with part-time special education, that support may end once they have acquired the typical age-level skills, and this requires systematic progress monitoring. In the present case, SEN teachers provided support over a longer period to students with more severe or comorbid difficulties while students who reached their classmates' typical skill level were supported for a shorter time.

4.4. Limitations

The present study has some limitations that should be addressed in future research. First, this study was conducted in educational settings within a single European country, and Finland's education system differs in a number of respects from other countries. For example, Finnish children enter first grade at seven years of age, which is 1 or 2 years later than in many other countries. Additionally, the provision of part-time special education is not based on any formal diagnosis. It follows that our results cannot necessarily be generalized to other learning contexts. Second, this study focused on duration rather than extent of support, which has been investigated using the same follow-up data from Grades 1 and 2 in a recent study by Holopainen et al. (2017). Third, we did not address the methods used by SEN teachers to identify support needs or to provide support, although teachers' ability and their approach to identifying reading difficulties in Grades 1 and 6 have been explored in recent studies using the same follow-up data (Virinkoski et al., 2018; Virinkoski et al., 2020). To deepen understanding of the relationship between part-time special education and skills development, it would be useful to incorporate earlier evidence. Finally, all of the achievement tests used in this study were group-administered, which can be more difficult to monitor and control in terms of ensuring that students correctly understand and are properly focused on the task.

4.5. Conclusions and implications for practice

This study examined the efficacy of part-time special education in Finland as currently implemented across Grades 1 to 4. Despite the scarcity of assessment tools, variations in assessment practices, and the absence of standardized follow-up procedures, special education resources in Finland seem to be used flexibly and in many cases align well with individual needs. This view is supported by the ESS group's ability to catch up with their classmates and the decisions taken by SEN teachers to confine extra support for these students to one or two years. A recent comparative study of reading acquisition rates in Norway and Finland by Solheim et al. (2020) confirmed that early identification followed by intensive support is the most effective solution to literacy learning difficulties. Targeted early intervention for students who need it most requires close attention to how students respond to instruction in the first months of school. This ensures that students whose progress is slower than expected can receive more intensive support in small groups while others continue with mainstream classroom instruction (Solheim et al., 2020). To provide targeted support, learning difficulties must be assessed using screening instruments that are sensitive to national educational context, including issues such as orthography (Solheim et al., 2020).

The present findings confirm that students who experience persistent difficulties in reading fluency in tandem with other learning-related difficulties typically receive part-time special education for several years. Unfortunately, it seems clear that this support does not help these students to catch up with their peers in terms of academic skills. Although the Finnish education system allocates significant resources to early needs-based support of this kind, support decisions generally depend on teachers' observations (Virinkoski et al., 2018; Virinkoski et al., 2020), and it appears that SEN teachers may lack both the time

and the competence to help students with multiple learning challenges (see Cancio et al., 2018; Langher et al., 2017). It seems likely that more effective and individually targeted methods of assessment and support are needed for these students, but further research is needed to clarify this issue.

In conclusion, for students with mild or limited learning difficulties, shorter support periods of part-time special education seem adequate to enable them to catch up with their peers. However, for students with persistent and multiple overlapping learning challenges, even the longer support currently offered seems insufficient. As we did not consider the nature and extent of that longer support, it remains a possibility that even individually tailored and carefully targeted support may not help those students.

References

- Aro, M. (2017). Learning to read Finnish. In L. Verhoeven, & C. Perfetti (Eds.), Learning to Read across Languages and Writing Systems (pp. 416–436). Cambridge University Press.
- Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, 96, 699–713. https://doi.org/10.1037/0022-0663.96.4.699.
- Psychology, 96, 699–713. https://doi.org/10.1037/0022-0663.96.4.699.

 Aunola, K., & Räsänen, P. (2007). Arithmetic tests. [Laskutaito testit]. Unpublished test material. University of Jyväskylä.
- Baird, G. (2008). Assessment and investigation of children with developmental language disorder. In C. F. Norbury, J. B. Tomblin, & D. V. M. Bishop (Eds.), *Understanding developmental language disorders*. From theory to practice (pp. 1–22). Psychology Press.
- Björn, P., Aro, M., Koponen, T. K., Fuchs, L. S., & Fuchs, D. H. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learning Disability Quarterly*, 39(1), 58–66. https://doi.org/10.1177/0731948715594787.
- Bottge, B. A., Cohen, A. S., & Choi, H. J. (2018). Comparisons of mathematics intervention effects in resource and inclusive classrooms. *Exceptional Children*, 84(2), 197–212. https://doi.org/10.1177/0014402917736854.
- Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46(1), 3–18. https://doi.org/10.1111/j.1469-7610.2005.00374 x
- Cabbage, K. L., Farquharson, K., Iuzzini-Seigel, J., Zuk, J., & Hogan, T. P. (2018). Exploring the overlap between dyslexia and speech sound production deficits. *Language, Speech, and Hearing Services in Schools, 49*, 774–786. https://doi.org/ 10.1044/2018 ISHSS-DYSIC-18-0008.
- Cancio, E. J., Larsen, R., Mathur, S. R., Estes, M. B., Johns, B., & Chang, M. (2018). Special education teacher stress: Coping strategies. *Education and Treatment of Children*, 41(4), 457–481. https://doi.org/10.1353/etc.2018.0025.
- Child, A. E., Cirino, P. T., Fletcher, J. M., Willcutt, E. G., & Fuchs, L. S. (2019). A cognitive dimensional approach to understanding shared and unique contributions to reading, math, and attention skills. *Journal of Learning Disabilities*, 52, 15–30. https://doi.org/10.1177/0022219418775115.
- Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. Contemporary Educational Psychology, 54, 99, 111. https://doi.org/10.1016/j.jcadpsyth.2018.06.002
- 99–111. https://doi.org/10.1016/j.cedpsych.2018.06.002.
 Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities*, 48(2), 156–175. https://doi.org/10.1177/0022219413494239.
- Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159. https://doi. org/10.1037/0033-2909.112.1.155.
- Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities*, 45(3), 232–243. https://doi.org/10.1177/0022219412442155.
- Denton, C. A., Fletcher, J. M., Anthony, J. L., & Francis, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities*, 39(5), 447–466. https://doi.org/10.1177/ 00222194060390050601.
- DuPaul, G. J., Gormley, M. J., & Laracy, S. D. (2013). Comorbidity of LD and ADHD: Implications of DSM-5 for assessment and treatment. *Journal of Learning Disabilities*, 46(1), 43–51. https://doi.org/10.1177/0022219412464351.
- 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. https://doi.org/10.1037/ a0037121.
- European Commission. (2013). Eurostat regional yearbook 2013: Education. http://ec.europa.eu/eurostat/en/web/products-statistical-books/-/KS-HA-13-001-04.
- Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problemsolving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities*, 37(4), 293–306. https://doi.org/10.1177/00222194040370040201.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A fiveyear prospective study. *Journal of Educational Psychology*, 104(1), 206–223. https:// doi.org/10.1037/a0025398.

- Griffiths, Y., & Stuart, M. (2013). Reviewing evidence-based practice for pupils with dyslexia and literacy difficulties. *Journal of Research in Reading*, 36(1), 96–116. https://doi.org/10.1111/j.1467-9817.2011.01495.x.
- https://doi.org/10.1111/j.1467-9817.2011.01495.x.

 Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., Denton, C. A., & Taylor, P. (2013).

 Anxiety and inattention as predictors of achievement in early elementary school children. *Anxiety, Stress & Coping, 26*(4), 391–410. https://doi.org/10.1080/10615806.2012.691969.
- Holopainen, L., & Hakkarainen, A. (2019). Longitudinal effects of reading and/or mathematical difficulties: The role of special education in graduation from upper secondary education. *Journal of Learning Disabilities*, 52(6), 456–467. https://doi. org/10.1177/0022219419865485.
- Holopainen, L. K., Kiuru, N. H., Mäkihonko, M. K., & Lerkkanen, M.-K. (2017). The role of part-time special education supporting students with reading and spelling difficulties from grade 1 to grade 2 in Finland. European Journal of Special Needs Education, 33(3), 316–333. https://doi.org/10.1080/08856257.2017.1312798.
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. Reading & Writing Quarterly, 25(1), 4–32. https://doi.org/10.1088/10.573560802491.208
- Ise, E., Blomert, L., Bertrand, D., Faísca, L., Puolakanaho, A., Saine, N. L., ... Schulte-Körne, G. (2011). Support systems for poor readers: Empirical data from six EU member states. *Journal of Learning Disabilities*, 44(3), 228–245. https://doi.org/10.1177/0022319410374235
- Isoaho, P., Kauppila, T., & Launonen, K. (2016). Specific language impairment (SLI) and reading development in early school years. Child Language Teaching and Therapy, 32, 147–157. https://doi.org/10.1177/0265659015601165.
- Jahnukainen, M., & Itkonen, T. (2016). Tired intervention: History and trends in Finland and the United States. European Journal of Special Needs Education, 31(1), 140–150. https://doi.org/10.1080/08856257.2015.1108042.
- Jordan, J.-A., Mulhern, G., & Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *Journal of Experimental Child Psychology*, 103, 455–468. https://doi.org/10.1016/j. iecp. 2009.01.011
- Kikas, E., Silinskas, G., & Soodla, P. (2015). The effects of children's reading skills and interest on teacher perceptions of children's skills and individualized support. *International Journal of Behavioral Development*, 39(5), 402–412. https://doi.org/ 10.1177/0165025415573641.
- Koponen, T., Aro, M., Poikkeus, A.-M., Niemi, P., Lerkkanen, M.-K., & Nurmi, J.-E. (2018). Comorbid fluency difficulties in reading and math: Longitudinal stability across early grades. Exceptional Children, 84(3), 298–311. https://doi.org/10.1177/ 0014402918756269
- Korhonen, J., Linnanmäki, K., & Aunio, P. (2014). Learning difficulties, academic well-being and educational dropout: A person-centred approach. *Learning and Individual Differences*, 31, 1–10. https://doi.org/10.1016/j.lindif.2013.12.011.
- Korpipää, H., Koponen, T., Aro, M., Tolvanen, A., Aunola, K., Poikkeus, A.-M., ... Nurmi, J.-E. (2017). Covariation between reading and arithmetic skills from grade 1 to grade 7. Contemporary Educational Psychology, 51, 131–140. https://doi.org/ 10.1016/j.cedpsych.2017.06.005.
- Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *Journal of Child Psychology and Psychiatry*, 51, 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.
- Langher, V., Caputo, A., & Ricci, M. A. (2017). The potential role of perceived support for reduction of special education teachers' burnout. *International Journal of Educational Psychology*, 6(2), 120–147. https://doi.org/10.17583/ijep.2017.2126.
- Lappalainen, K., & Hotulainen, R. (2012). Participation in part-time special education and its correlation with the educational paths, self-concepts and strengths of young adults. British Journal of Special Education, 39(4), 185–193. https://doi.org/10.1111/ 1467-8758 12002
- Lemons, C. J., Vaughn, S., Wexler, J., Kearns, D. M., & Sinclair, A. C. (2018). Envisioning an improved continuum of special education services for students with learning disabilities: Considering intervention intensity. Learning Disabilities Research & Practice, 33(3), 131–143. https://doi.org/10.1111/jdrp.12173.
- Lerkkanen, M.-K. (2007). The beginning phases of reading literacy instruction in Finland. In P. Linnakylä, & I. Arffman (Eds.), Finnish reading literacy. When quality and equity meet (pp. 155–174). University of Jyväskylä, Institute of Educational Research.
- Lerkkanen, M.-K., Ahonen, T., & Poikkeus, A.-M. (2011). The development of reading skills and motivation and identification of risk at school entry. In M. Veisson, E. Hujala, P. K. Smith, M. Waniganayake, & E. Kikas (Eds.), Global perspectives in early childhood education: Diversity, challenges and possibilities (pp. 237–258). Peter Lang.
- Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, E., Siekkinen, M., & Nurmi, J.-E. (2006-2016). The first steps study [Alkuportaat]. Finland: University of Jyväskylä, University of Turku, and University of Eastern Finland. https://www.jyu.fi/alkuportaat.
- Lerkkanen, M.-K., & Poikkeus, A.-M. (2009). Reading comprehension: sentences.
- Unpublished measure. *Jyväskylä: University of Jyväskylä.*Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2004). The developmental dynamics of literacy skills during the first grade. *Educationa Psychology*, 24(6), 793–810.
- Lindeman, J. (1998). ALLU—Reading test for primary school [Ala-asteen lukutesti ALLU].

 Centre for Learning Research, University of Turku.
- Maehler, C., & Schuchardt, K. (2016). Working memory in children with specific learning disorders and/or attention deficits. *Learning and Individual Differences*, 49, 341–347. https://doi.org/10.1016/j.lindif.2016.05.007.

- Mancilla-Martinez, J., & Lesaux, N. K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, 102(3), 701–711. https://doi.org/10.1037/a0019135.
 Massetti, G., M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., &
- Massetti, G., M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for Attention-Deficit/Hyperactivity Disorder at 4–6 years of age. *Journal of Abnormal Psychology*, 36, 399–410. https://doi.org/10.1007/s10802-007
- Miciak, J., Roberts, G., Taylor, W. P., Solis, M., Ahmed, Y., Vaughn, S., & Fletcher, J. M. (2017). The effects of one versus two years of intensive reading intervention implemented with late elementary struggling readers. *Learning Disabilities Research and Practice*, 33(1), 24–36. https://doi.org/10.1111/ldrp.12159.
- Moll, K., Gangl, M., Banfi, C., Schulte-Körne, G., & Landerl, K. (2020). Stability of deficits in reading fluency and/or spelling. Scientific Studies of Reading, 24(3), 1–11. https://doi.org/10.1080/10888438.2019.1659277.
- Moll, K., Landerl, K., Snowling, M. J., & Schulte-Körne, G. (2019). Understanding comorbidity of learning disorders: Task-dependent estimates of prevalence. *Journal* of Child Psychology and Psychiatry, 60(3), 286–294. https://doi.org/10.1111/ jcnp.12965
- Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, 42, 306–321. https://doi.org/10.1177/0022219408331037.
- Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics. Who falls increasingly behind? *Journal of Learning Disabilities*, 44(5), 472–488. https://doi.org/10.1177/0022219411414010.
- Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, 51(6), 523–539. https://doi. org/10.1177/0022919417714773
- Official Statistics of Finland. (2018). Special education [e-publication]. Statistics Finland. http://www.stat.fi/til/erop/tau/en.html.
- Parhiala, P., Torppa, M., Vasalampi, K., Eklund, K., Poikkeus, A.-M., & Aro, T. (2018). Profiles of school motivation and emotional well-being among adolescents: Associations with math and reading performance. *Learning and Individual Differences*, 61, 196–204. https://doi.org/10.1016/j.lindif.2017.12.003.
- Peterson, R. L., Boada, R., McGrath, L. M., Willcutt, E. G., Olson, R. K., & Pennington, B. F. (2017). Cognitive prediction of reading, math, and attention: Shared and unique influences. *Journal of Learning Disabilities*, 50(4), 408–421. https://doi.org/10.1177/0022219415618500.
- Prochnow, J. E., Tunmer, W. E., & Chapman, J. W. (2013). A longitudinal investigation of the influence of literacy-related skills, reading self-perceptions, and inattentive behaviours on the development of literacy learning difficulties. *International Journal* of Disability, Development and Education, 60(3), 185–207. https://doi.org/10.1080/ 1034013V 2013 813188
- Pulkkinen, J., Räikkönen, E., Jahnukainen, M., & Pirttimaa, R. (2019). How do educational reforms change the share of students in special education? Trends in special education in Finland. European Educational Research Journal, 19(4), 364–384. https://doi.org/10.1177/1474904119892734.
- Share, D. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. Psychological Bulletin, 134(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584.
- Skibbe, L. E., Grimm, K. J., Stanton-Chapman, T. L., Justice, L. M., Pence, K. L., & Bowles, R. P. (2008). Reading trajectories of children with language difficulties from preschool through fifth grade. Language, Speech & Hearing Services in Schools, 39(4), 475-486. https://doi.org/10.1044/0161-1461(2008/07-0016).
- 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/ 10.1027/bul0000037
- Solheim, O. J., Torppa, M., Uppstad, P. H., & Lerkkanen, M.-K. (2020). Screening for slow reading acquisition in Norway and Finland a quest for context specific predictors.

- Scandinavian Journal of Educational Research. https://doi.org/10.1080/00313831.2020.1739130.
- Soodla, P., Tammik, V., & Kikas, E. (2019). Is part-time special education beneficial for children at risk for reading difficulties? An example from Estonia. *Dyslexia*, 1–25. https://doi.org/10.1002/dys.1643.
- Spanoudis, G. C., Papadopoulos, T. C., & Spyrou, S. (2019). Specific language impairment and reading disability: Categorical distinction or continuum? *Journal of Learning Disabilities*, 52(1), 3–14. https://doi.org/10.1177/0022219418775111.
- Takala, M., & Ahl, A. (2014). Special education in Swedish and Finnish schools: Seeing the forest or the trees? *British Journal of Learning Disabilities*, 41(1), 59–81. https://doi.org/10.1111/1467-8578.12049.
- Takala, R., Pirttimaa, R., & Törmänen, M. (2009). Inclusive special education: The role of special education teachers in Finland. British Journal of Special Education, 36(3), 162–172. https://doi.org/10.1111/j.1467-8578.2009.00432.x.
 Torgesen, J., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling
- Torgesen, J., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling readers. In S. J. Samuels, & A. Farstrup (Eds.), Reading fluency: The forgotten dimension of reading success (pp. 130–158). International Reading Association.
- Torppa, M., Eklund, K., van Bergen, E., & Lyytinen, H. (2015). Late-emerging and resolving dyslexia: 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., Niemi, P., Lerkkanen, M.-K., & Poikkeus, A.-M. (2017). The precursors of double dissociation between reading and spelling in a transparent orthography. *Annals of Dyslexia*, 67(1), 42–62. https://doi.org/10.1007/s11881-016-0131-5.
- 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.
- Van der Lely, H. K. J., & Marshall, C. R. (2010). Assessing component language deficits in the early detection of reading difficulty risk. *Journal of Learning Disabilities*, 43(4), 357–368. https://doi.org/10.1177/0022219410369078.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*, 18(3), 137–146. https://doi.org/10.1111/1540-5826.00070.
 Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with
- Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. Learning Disabilities Research & Practice, 29 (2), 46, 53
- Virinkoski, R., Lerkkanen, M.-K., Eklund, K., & Aro, M. (2020). Special education teachers' identification of students' reading difficulties in Grade 6. Scandinavian Journal of Educational Research. https://doi.org/10.1080/00313831.2020.1833241.
- Virinkoski, R., Lerkkanen, M.-K., Holopainen, L., Eklund, K., & Aro, M. (2018). Teachers' ability to identify children at early risk for reading difficulties in Grade 1. Early Childhood Education Journal, 46(5), 497–509. https://doi.org/10.1007/s10643-017-0883-5
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2009). TOSREC: Test of sentence reading efficiency and comprehension. Pro-Ed.
 Wanzek, J., Al Otaiba, S., & Petscher, Y. (2014). Oral reading fluency development for
- Wanzek, J., Al Otaiba, S., & Petscher, Y. (2014). Oral reading fluency development for children with emotional disturbance or learning disabilities. *Exceptional Children*, 80 (2), 187–204. https://doi.org/10.1177/001440291408000204.
 Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., &
- Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disabilities*, 46(6), 500–516. https://doi.org/10.1177/0022210413477476
- Zhang, X., Räsänen, P., Koponen, T., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2020). Early cognitive precursors of children's mathematics learning disability and persistent low achievement: A five-year longitudinal study. *Child Development*, 91(1), 7–27. https://doi.org/10.1111/cdev.13123.