

**JYX**



JYVÄSKYLÄN YLIOPISTO  
UNIVERSITY OF JYVÄSKYLÄ

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Virinkoski, Riitta; Lerkkanen, Marja-Kristiina; Eklund, Kenneth; Aro, Mikko

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

**Year:** 2022

**Version:** Published version

**Copyright:** © 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Virinkoski, R., Lerkkanen, M.-K., Eklund, K., & Aro, M. (2022). 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>

# 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](https://doi.org/10.1080/00313831.2020.1833241)

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



© 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 [↗](#)



Article views: 1358



View related articles [↗](#)







View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

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

Riitta Virinkoski <sup>a,b</sup>, Marja-Kristiina Lerkkanen <sup>b</sup>, Kenneth Eklund <sup>c</sup> and Mikko Aro <sup>d</sup>

<sup>a</sup>Open University, University of Jyväskylä, Jyväskylä, Finland; <sup>b</sup>Department of Teacher Education, University of Jyväskylä, Jyväskylä, Finland; <sup>c</sup>Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland; <sup>d</sup>Department 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 long-term 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

**CONTACT** Riitta Virinkoski  riitta.e.virinkoski@jyu.fi  Open University, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland Department of Teacher Education, University of Jyväskylä, P.O. Box 35, 40014 Jyväskylä, Finland

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 pre-reading 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 systematically 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 sub-skills 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 (Lerikkanen 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 meta-analyses 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 15<sup>th</sup> percentile in at least two of the measured skill areas or if they scored at or below the 15<sup>th</sup> 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 group-administered 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<sup>th</sup> 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<sup>th</sup> 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<sup>th</sup> 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,



**Table 1.** Descriptive statistics of SEN Teachers' ( $n = 29$ ) Reading Assessment 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 <sup>nd</sup> most important n (%)	The 3 <sup>rd</sup> most important n (%)	The 4 <sup>th</sup> 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 ( <i>n</i> = 55)	63	69	18 (10)	22 (12)	49 (27)	11 (6)
Reading comprehension ( <i>n</i> = 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<sup>th</sup> 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 ( $\chi^2(1) = 4.72, p = .030$ ). 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	B	S.E.	OR	<i>p</i>	R <sup>2</sup>
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<sup>2</sup> = 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 low-performing 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 be 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, decision-making 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 16<sup>th</sup> 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 16<sup>th</sup> 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 Nyssönen 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](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>
- Coladarcì, 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](https://doi.org/10.1177/0022466903037003080_1)
- Dixon, 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., Mäkihönko, 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., Faisca, 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 Aus- und 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 [Alkuportaati], 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>