



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

Author(s): Juhkam, Maris; Soodla, Piret; Aro, Mikko

Title: How accurate are teachers and support specialists when judging students' literacy skills? Special educational service as an external factor influencing judgements

Year: 2022

Version: Published version

Copyright: © 2022 The Authors. *Dyslexia* published by John Wiley & Sons Ltd.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Juhkam, M., Soodla, P., & Aro, M. (2022). How accurate are teachers and support specialists when judging students' literacy skills? Special educational service as an external factor influencing judgements. *Dyslexia*, 28(4), 378-396. <https://doi.org/10.1002/dys.1725>

How accurate are teachers and support specialists when judging students' literacy skills? Special educational service as an external factor influencing judgements

Maris Juhkam¹  | Piret Soodla¹  | Mikko Aro² 

¹School of Educational Sciences, Tallinn University, Tallinn, Estonia

²Department of Education, University of Jyväskylä, Jyväskylä, Finland

Correspondence

Maris Juhkam, School of Educational Sciences, Tallinn University, Narva mnt 25, Tallinn 10120, Estonia.
Email: maris.juhkam@tlu.ee

Funding information

EEA Financial Mechanism 2014-2021 Higher Education in Baltic Research Programme through the Project "Developing and Enhancing the Teaching Quality of Inclusive Education Curriculum", Grant/Award Number: 36.1-3.4/289

The aim of the study was to examine the accuracy of Estonian teachers' and support specialists' judgements of students' spelling skills and reading fluency and to investigate the provision of special education services to students as a factor influencing teachers' judgements. The sample included 11 classroom teachers, 8 support specialists, and 187 third-grade students. The judgements were collected using scales, and students' literacy skills were assessed using group and individual tests. The results indicated that judgements of reading fluency were less accurate than those of spelling skills. In addition, the provision of special education services influenced teachers' judgements, directing them to identify students in need of help, even if the teacher did not initially assign the student to the low-skilled group. Unexpectedly, teachers' judgements of the skills of students receiving special education services were slightly more accurate than support specialists' judgements.

KEY WORDS

judgements, literacy, special education service, support specialists, teachers

Practitioner Points

- Teachers' judgements of students' fluency were less accurate in reading than in spelling.
- A student's "special educational status" may influence teachers' judgements.
- The results suggest that collaboration between teachers and support specialists when assessing students' skill levels may be insufficient.

1 | INTRODUCTION

Teachers' assessments of students' progress in learning is a primary source for identifying students in difficulty as early as possible (Bailey & Drummond, 2006; Thiede et al., 2018). Therefore, assessing students' skills is one of teachers' main tasks when planning an effective learning process (Clark & Peterson, 1986; Shavelson & Stern, 1981) and, when necessary, assigning special educational support (VanDerHeyden, Witt, & Naquin, 2003). Moreover, educational decisions regarding a student's further development and support are more effective if decisions are made based on purposefully collected data on the student's skills—namely, data-based decisions (Earl & Louis, 2013). Data use interventions for teachers have shown that teachers can successfully use data collected on students' skills to support students in overcoming difficulties (Poortman & Schildkamp, 2016). However, it has been found that in daily work teachers make relatively few data-based decisions, and instead tend to make decisions based on intuition (Vanlommel, Van Gasse, Vanhoof, & Van Petegem, 2017).

Although in general, teachers' judgements of students' academic skills have been extensively studied (Hoge & Coladarci, 1989; Südkamp, Kaiser, & Möller, 2012), there are still bottlenecks that this study seeks to address. First, recent research has focused on teachers' judgements of students' reading fluency (Begeny, Eckert, Montarello, & Storie, 2008; Feinberg & Shapiro, 2003) and, more broadly, their reading skills (Bates & Nettelbeck, 2001; Feinberg & Shapiro, 2009; Soodla & Kikas, 2010), but fewer studies have been conducted in the field of writing skills (Ritchey, Coker, & Jackson, 2015; Wheadon, Barmby, Christodoulou, & Henderson, 2020). The present study focuses on both reading fluency and elementary spelling. Second, previous studies of teachers' judgements have mostly used scales without defined midpoints, only endpoints, and often in a very general form, such as a 5-point Likert-type scale designated as 5 (very successful) and 1 (very low success) (Begeny et al., 2008; Feinberg & Shapiro, 2003). In this study, we used scales where each point on the scale contained a brief description of the skill level. The idea was to ensure greater objectivity in making judgements and, thus, obtain information on the accuracy of judgements. Third, whereas recent research has focused mainly on teachers' judgements (Begeny et al., 2008; Feinberg & Shapiro, 2003; Soodla & Kikas, 2010), we examined the judgement accuracy of classroom teachers as well as support specialists (special education teachers, speech therapists, remedial teachers), both of whom are responsible for teaching and supporting students' literacy skills during elementary school years.

1.1 | Judgements of students' reading fluency and spelling skills

The correlations between teachers' judgements and students' academic performance according to test results have been found to be predominantly moderate (Hoge & Coladarci, 1989; Südkamp et al., 2012). Focusing in more detail on reading fluency, the correlations also tend to be moderate (Eckert, Dunn, Codding, Begeny, & Kleinmann, 2006; Feinberg & Shapiro, 2003). However, several previous studies have shown that teachers tend to overestimate weak readers' skills (Bates & Nettelbeck, 2001; Begeny et al., 2008; Feinberg & Shapiro, 2009; Soodla & Kikas, 2010). Regarding students' spelling skills, one study found that teachers were rather accurate at identifying at-risk preschool

children who may experience spelling difficulties in school (Kolne, Gonnerman, Marquis, Royle, & Rvachew, 2016), but no relevant research was found in the context of assessing students' spelling skills at school. The accuracy of support specialists' judgements has been particularly understudied, with few relevant studies being conducted in Finland (Virinkoski, Lerkkanen, Holopainen, Eklund & Aro, 2017; Virinkoski, Lerkkanen, Eklund & Aro, 2020). One of these studies found that special educators' judgements were more accurate than those of classroom teachers, but identifying children with weak phonological skills were still challenging (Virinkoski et al., 2017).

Studies have also revealed that several external factors may affect the judgements. Johansson, Myrberg, and Rosen (2012) found that the longer the teaching period was, the more accurate the judgements were. However, a longitudinal study in which teachers were asked to rate students' reading skills from first to third grades found that teachers' ratings were strongly influenced by students' past performance (Hecht & Greenfield, 2002). Thus, students' earlier performance may also influence teachers, causing them to give incorrect judgements. Other studies have revealed that teachers tend to underestimate the academic skills of students with special educational needs or students who have previously needed special help (Campbell, 2015; Hurwitz, Elliott, & Braden, 2007; Soodla & Kikas, 2010) and their expectations for the further academic success of such students are lower (Vlachou, Eleftheriadou, & Metallidou, 2014; Woodcock & Hitches, 2017). Furthermore, Kikas, Soodla, and Mägi (2018) found that the general level of the class may influence teachers' judgements: When classmates have higher reading and math skills, teachers tended to assess the individual students' respective skills rather low and vice versa.

1.2 | Learning and teaching to read and spell in Estonian schools

Learning to read and spell depends on the orthography of the language. In languages with transparent orthography (e.g., Finnish, Italian, German), where there is a consistent grapheme–phoneme correspondence, basic reading skills are acquired significantly faster than in languages with deeper orthography (e.g., English), where grapheme–phoneme correspondence is more inconsistent (Seymour, Aro & Erskine, 2003; van Daal & Wass, 2017). Moreover, in transparent orthographies, reading speed is more of a concern than reading accuracy (Landerl & Wimmer, 2008). The Estonian language has a transparent orthography (Viise, Richards, & Pandis, 2011); therefore, learning to read is usually fast and easy for children. Most Estonian children entering school are able to read and, by the end of first grade, the accuracy and fluency of their reading are already rather high (Soodla et al., 2015). The peculiarity of the Estonian language and orthography is the need for distinguishing sound length in three ways: short, long, and overlong (Lippus, Pajusalu, & Allik, 2009). Because these three lengths are difficult to perceive and distinguish, students often make spelling errors, particularly when marking the length of plosive consonants (*p*, *k*, and *t*; e.g., in the word *häppavad* "jump," where "*p*" is overlong, students mark "*p*" as long—that is, with one letter—and write *häpavad*) (Karlep, 2000; Lippus et al., 2009; Viise et al., 2011). In addition, long word forms, which often contain many suffixes, are typical in Estonian and cause difficulties in reading and writing (Hint, 1998; Soodla, Vija & Pajusalu, 2013).

In Estonia, students are required to complete nine grades. Schools that provide this compulsory education are called basic schools (see the Ministry of Education and Research website for more information: <https://www.hm.ee/en>). Basic schools must follow the National Curriculum for Basic Schools (Government of the Estonian Republic, 2011), which determines which skills a student should acquire by the end of specific grades (i.e., by the end of third, sixth, and ninth grades). According to the curriculum, students should be able to read texts fluently with good comprehension and apply basic spelling rules when writing by the end of third grade. However, data show that about 22% of students in basic school have special educational needs, most of whom need support in developing reading and writing skills (Kallaste, 2016). If a student has learning difficulties, the necessary support will be provided at school. Estonian schools have a three-level system of support for students with difficulty: general support, enhanced support, and special support (Government of the Estonian

Republic, 2010). Assigning enhanced or special support to the student requires the involvement of an out-of-school counseling team and, in most cases, a diagnosis of the student. Whether a student needs general support is an in-school pedagogical decision and does not require a formal diagnosis of learning disabilities. General support is usually arranged as a pull-out service, and the frequency of support lessons depends on the individual needs, but students typically attend such sessions twice a week (Soodla, Tammik, & Kikas, 2020). In addition to special educators, speech therapists and remedial teachers work in Estonian schools and support students with reading and writing difficulties (Padrik & Kikas, 2007). Classroom teachers, special education teachers, and speech therapists are required to have a master's degree to work at schools in Estonia. Both special education teachers and speech therapists learn to support the development of reading and writing skills of students with learning difficulties. Remedial teachers are required to have a master's degree in pedagogy (e.g., in classroom teaching or subject teaching) as well as complete in-service training on how to support students with learning difficulties (especially students with reading and writing difficulties, as these are the most common difficulties for students).

The concern is that there is a lack of reliable assessment tools in Estonian schools and no standardized tests for assessing literacy skills (such as fluency in reading and spelling skills) are available. Therefore, teachers make their own assessment tools. For example, teachers often use dictation tasks to assess students' spelling skills or reading texts selected or adapted from school books to assess reading skills. Teachers also gather information through daily observations to evaluate students' skills and guide pedagogical decisions. If the classroom teacher concludes from the information gathered using self-made tools and observations that the student is having difficulty, the teacher will involve a support specialist. Although support specialists also lack reliable assessment tools, their professional training means they can design more detailed assessment tools of varying difficulty to better identify the student's skill level and area of proximal development.

1.3 | Present study

The present study had two purposes: to examine the accuracy of teachers' and support specialists' judgements of third-grade students' spelling skills and reading fluency and to analyse the impact of the provision of special educational services on teachers' judgements. To achieve these purposes, we developed the following hypotheses:

Hypothesis 1. As spelling skills are "more visible" than reading fluency, we expect teachers' judgements of students' spelling skills to be more accurate than their judgements of students' reading fluency.

Hypothesis 2. As previous research has shown (Bates & Nettelbeck, 2001; Feinberg & Shapiro, 2009), we expect that teachers tend to overestimate the skill level of low-skilled readers. Although no specific studies on the assessment of spelling skills were found, based on the general tendency for teachers to overestimate the performance level of students with lower academic skills, we expect that the performance of weak spellers is also overestimated.

Hypothesis 3. Teachers tend to underestimate the skills of students who receive part-time special educational support (Campbell, 2015; Hurwitz et al., 2007; Soodla & Kikas, 2010).

Hypothesis 4. Support specialists assess the literacy skills of students receiving part-time special education services more accurately than do teachers (Virinkoski et al., 2017).

2 | METHOD

2.1 | Participants

The sample included eight general education schools (i.e., schools teaching according to the common national curriculum) and 11 third-grade classes. The language of instruction of all schools was Estonian. The average class size was 17.46 students ($SD = 6.79$). In total, 187 third graders participated, and 34 of them received part-time special educational services (i.e., they participated in special education lessons at school). The average age of the students was 9.83 years ($SD = 0.37$). Four students (2% of the sample) spoke two home languages according to their parents: Estonian along with Russian, Finnish, or English. The Ethics Committee of the National Institute for Health Development approved the study, and all students and their parents gave informed consent to participate (93% of those initially contacted).

The sample also consisted of eight support specialists and 11 classroom teachers; all were female. Classroom teachers had taught the participating students for 3 years. The teachers' overall work experience was lower than that of the average Estonian teacher, which is about 23 years (OECD, 2019). The support specialists had taught the participating students for 2–3 years as well as during the data collection period. Data on participants' educational level and teaching experience are presented in Table 1.

2.2 | Materials and procedure

2.2.1 | Tools for collecting teachers' and support specialists' judgements

Teachers and support specialists assessed students' skills in January 2020, during the morning of the first day of an in-service training session. The training aimed to provide teachers with knowledge on how to teach and guide students in acquiring and using reading comprehension strategies. Data were collected on paper. Teachers and support specialists were asked to think about their students' reading fluency and spelling skills in general and rate their students' spelling and reading skills using the descriptions on the scales. The scale point descriptions were compiled by the first and second authors based on the national curriculum and previous knowledge about the development of Estonian children's reading and spelling skills (Karlep, 2000; Soodla et al., 2015; Viise et al., 2011). The teachers rated each student while the support specialists rated only the students they taught (in special education lessons). The first and second authors remained in the same room as the teachers and support specialists during data collection and asked them not to cooperate during the assessments. Throughout the procedure, the researchers ensured that participants gave their assessments of students' skills individually.

TABLE 1 Participants' education level and classroom experience

Participants	Education	Classroom experience
Classroom teachers ($n = 11$)	Master's degree in classroom teaching: 7 teachers Bachelor's degree in preschool pedagogy: 2 teachers Master's degree in classroom teaching in process: 1 teacher Level of education not indicated: 1 teacher	1–5 years: 8 teachers 6+ years: 3 teachers
Support specialists ($n = 8$)	Special educator with a master's degree: 5 support specialists Speech therapist with a master's degree: 1 support specialist Remedial teacher with a master's degree in philology: 1 support specialist Remedial teacher, no indication of level of education: 1 support specialist	6–10 years: 1 support specialist 11–15 years: 7 support specialists

The scale used to assess spelling included the following points: 4 = The student marks the duration of speech sounds mostly correctly when writing (rarely makes mistakes when indicating the duration of t, k, p) and correctly applies learned orthographic rules when writing; 3 = The student makes some mistakes when marking the duration of speech sounds (especially when indicating the duration of t, k, p, but rarely when noting the duration of other consonants) and sometimes makes mistakes when applying learned orthographic rules; 2 = The student often makes mistakes when marking the duration of speech sounds and often makes mistakes when applying learned orthographic rules; 1 = The student makes a lot of mistakes when marking the duration of speech sounds and applying learned orthographic rules.

The scale used to assess reading fluency included the following scale points: 4 = The student reads texts fluently (at the pace of normal speaking) and correctly; 3 = The student reads mostly fluently (at the pace of speaking), but is slower when reading long words, and reads mostly correctly; 2 = The pace of reading is slower than the pace of speaking, and the student struggles with longer words; the student sometimes makes mistakes when reading longer words; 1 = The rate of reading is slow, and the student struggles with longer words; the student often makes mistakes when reading words.

Once the scale-based judgements were given, the papers were immediately collected. We then asked teachers to indicate those students who were currently receiving part-time special education and specify in which of the four skill areas (oral speech, reading, writing, and math) the student needed additional support. They were instructed to choose all categories in which the student needed additional support, but the present analysis focused only on the students identified as needing support in reading, writing, or both. The teachers were also asked to nominate the students who, in their opinion, need additional help in some skill areas, but did not participate in special education lessons at school at that time for some reason. In such cases, a justification was also requested (e.g., the student receives support outside of school). When completing the questionnaire, the teachers could consult with support specialists.

2.2.2 | Tests to assess students' literacy skills

Students' spelling skills were assessed using two group-administered tests: dictated phrase writing test and sentence correction test. The tests contained words that required the duration of speech sounds to be spelled correctly. For example, students had to write the phrase *häppavad palliga* ("they) are jumping with a ball"; in the word *häppavad*, "p" is overlong and therefore has to be marked with two "p" letters while in *palliga* "l" is long sound and has to be marked with two "l" letters. The spelling tests also contained words requiring the application of orthographic rules (e.g., the student had to correct mistakes in the misspelled word *lillet* "flowers," which requires "d" instead of "t" to indicate plural in Estonian).

The phrase writing test contained 17 phrases (42 words), with each correctly written word being worth one point (Max = 42 points). The researcher read out the phrase and repeated it once, after which the students wrote the phrase on a sheet of paper. In the sentence correction test, students had to analyse 24 words (within 6 sentences), of which 21 contained a spelling error. Each misspelled word (21 words) that the student corrected as well as each correctly spelled word that the student left uncorrected (3 words) was worth one point (Max = 24 points). For the phrase writing test, the Cronbach's alpha was .98; for the sentence correction test, it was .94. The alpha across the two tests was .98. The material selected for the tests corresponded to the topics covered in the national curriculum. The task was not time limited, and the researchers ensured that all students completed the tests. As the correlation between the results of the phrase writing test and the results of the sentence correction test was strong ($r_s = .841, p < .01$), the results of the two tests were summed to give the final score of spelling (Max = 66 points). Spelling tests were experimental measures designed for the present study and were compiled by the first author.

Students' reading fluency was assessed by two individually administered tests: a connected narrative text reading test and a word-list reading test. Students were first asked to read aloud a connected text, with a time limit of

60 seconds. The reading was recorded, and the number of correctly read words was used as a score. The text about how the children fished, was selected from the children's book. The complexity of the text (e.g., the length of words and sentences) was comparable to the complexity of the texts in the third-grade school books. The text consisted of 32 sentences, with a mean sentence length of 6 words. The text had 197 words total, of which 39 (approximately 20%) were longer words (i.e., more than six characters). After reading the texts, students were asked to read aloud a word list containing 90 words (placed in three columns, 3x30 words) within 60 s. The words in the word list were not related to the text the students had read before. The words' structures became increasingly complex. The words in the first column were familiar to students and consisted of one to two syllables (three to seven characters; e.g., *maa* "land"; *ema* "mother"; *koer* "dog"; *reede* "Friday"; *traktor* "tractor"). The words in the second column consisted of three to five syllables (6–13 characters; e.g., *vaarikas* "raspberry"; *artistlik* "artistic"; *kirjutuslaud* "desk"; *kõrvaklapid* "headphones"). Some words in the second column may have been less familiar to students. The words in the third column consisted of three to eight syllables (8–20 characters; e.g., *klaverimängija* "piano player"; *filmiprodutsent* "film producer"; *indiaanlane* "Indian"; *arhitektuuribüroo* "architectural office"). Several words in the third column may have been less familiar to students. However, due to the time limit, most students did not reach the third column. The reading was recorded, and the number of correctly read words was used as a score. Given the strong correlation between the scores of the text reading and word-list reading tasks ($r_s = .857$, $p < .01$), the results of the two tasks (i.e., the number of correctly read words) were summed to give the final score of reading fluency. Reading tests were compiled by the second author.

Data collection from students was conducted from November 2019 to December 2019. All test administration was carried out at the students' school on a single school day, starting with the spelling tasks. Before testing, the researcher (the first author or a trained undergraduate student) explained the tasks to the students and went through some practice items.

2.3 | Data coding and analysis

To determine the correspondence between the judgements and the test results, the coding system was created. Cut-off points were determined based on the expected learning outcomes formulated in the national curriculum (Government of the Estonian Republic, 2011) and the joint decisions of the first and second authors and an expert teacher, all of whom had a master's degree in special education. This expert teacher was not involved in the study other than the determination of cut-off points.

As the material used in the spelling test corresponded to the national curriculum, to distinguish skilled writers, it was jointly agreed that the limit of errors should be minimal (up to three errors; i.e., at least 95% of words spelled correctly). A number of words in the spelling tests required indicating the duration of speech sounds, which could be difficult for many students, not only the ones with weakest performance; thus, it was decided that students in the average group could have up to 14 mistakes (i.e., spelled 80%–95% of words correctly). Thus, making more than 14 mistakes indicated that students' spelling skills did not correspond to what is expected in the curriculum. Based on the final scores, three groups were formed: better spellers (63–66 points), average spellers (52–62 points), and poorer spellers (fewer than 52 points). The group of better spellers included 65 students (35%), the average group had 85 students (45%), and the poorer group had 40 students (21%). Students in the average and better spellers groups spelled most of the words correctly; thus, their skill level can be considered to meet the curriculum requirements. Poorer spellers made many spelling mistakes and, thus, might be in need of additional support. Based on teachers' and support specialists' judgements, three groups were formed: rating 4 = better spellers; rating 3 = average spellers; ratings 2 and 1 = poorer spellers.

Students' reading fluency level was assessed based on a connected text reading test. The two experts (the first author and a graduate student in special education) trained together and then independently listened to recordings of students reading the connected text; they used the same scale as teachers and support specialists to assess

students' reading fluency. The overall Cohen's kappa between the two experts' judgements was .74; in the group of poorer readers, the Cohen's kappa was .88. The first author once again listened to the recordings for which the assessments differed and gave a final assessment of the level of fluency of students' reading. Based on the expert judgements, three groups were formed: rating 4 = better readers (64 students; 35%); rating 3 = average readers (82 students; 45%); ratings 2 and 1 = poorer readers (38 students; 21%). Three groups were also formed based on teachers' and support specialists' judgements: rating 4 = better readers; rating 3 = average readers; ratings 2 and 1 = poorer readers. The total score of the reading tests (i.e., the number of words read correctly in both the connected text and the word-list test) was used only to describe the correlation between students' performance and the judgements. The correlation between expert judgement and the number of words read correctly was .831 ($p < .01$).

In the data analysis, a Spearman correlation was used to describe the strength of the correlations between judgements and test results. To identify the statistical significance of consistencies between the variables, a Chi-square test was used. Fisher's exact test was performed if the Chi-square test conditions related to the sample size or the number of observed variables were not met. Cramer's V was used to find the effect size (Cohen, 1988). In order to find statistically significant associations between judgements and test results in the groups of poorer, average, and better-skilled students, a consistency table was compiled. Results were interpreted using adjusted residual values.

3 | RESULTS

3.1 | Correspondence of teachers' judgements and test results

First, the relationship between teachers' judgements on students' spelling skills and the test results were analysed. The mean score of the spelling test was 56.73 ($SD = 10.78$, Min = 11, Max = 66), and the correlation between judgements and test scores was .643 ($p < .01$). Next, the consistency between the groups formed based on the test results and the teachers' judgements were examined. A Chi-square test of independence showed a significant association between the categories of teachers' judgements and test results: $\chi^2(4, N = 187) = 99.99$, $p < .001$, $V = .52$. The correspondences between teachers' judgements and test results in assessing spelling are presented in Table 2.

TABLE 2 Correspondence between teachers' judgements and test results in assessing spelling skills

		Teachers' judgements			
Spelling test results		Poorer speller	Average speller	Better speller	Total (N = 187)
Poorer speller	f_o	33	7	0	40
	f_e	10.27	15.83	13.90	40
	AR	9.28	-3.22	-5.21	
Average speller	f_o	15	40	27	82
	f_e	21.05	32.45	28.50	82
	AR	-2.04	2.28	-.47	
Better speller	f_o	0	27	38	65
	f_e	16.68	25.72	22.59	65
	AR	-5.87	.40	4.97	
Total (N = 187)		48	74	65	187

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

The examination of adjusted residuals (AR) indicated that the observed frequency was random in the following categories: average speller–better speller and better speller–average speller. This suggests that teachers had difficulty distinguishing between students with average or better skills based on test results. According to the test results, 40 students belonged to the group of poorer spellers. In this subgroup, consistency between teachers' judgements and test results occurred in 33 cases (83%) while overestimation occurred in 7 cases (17%). Four of these seven students scored very poorly on the spelling test (made more than 20 mistakes; i.e., they misspelled more than 30% of the words).

Second, the relationship between teachers' judgements of students' reading fluency and the test results was analysed. The mean score of the reading tests was 120.41 ($SD = 29.13$, Min = 42, Max = 187), and the correlation between the judgements and the test score was .589, $p < .01$. A Chi-square test of independence showed a significant association between the categories of teachers' judgements and test results: $\chi^2(4, N = 184) = 66.42, p < .001$, $V = .42$. The correspondences between teachers' judgements and test results in assessing reading fluency are presented in Table 3.

The examination of AR indicated that the observed frequency was random in the following categories: average reader–poorer reader, poorer reader–average reader, and better reader–average reader. This suggests that teachers had difficulty distinguishing between students in the poorer and average groups based on the test results as well as those who belonged to the average and better group. According to the test results, 38 students belonged to the group of poorer readers. In this subgroup, consistency between teachers' judgements and test results occurred in 24 cases (63%). Overestimation occurred in 14 cases (37%); in 10 cases the student was assigned to the group of average readers and in 4 cases to the group of better readers.

3.2 | Teachers' judgements and students' participation in special education lessons

Based on scale judgements given by teachers, 57 students belonged to the poorer group of spellers or readers or both. Of these students, 33 (58%) participated in special education lessons focused on developing literacy skills while 24 (42%) did not. Teachers' judgements and information on students' participation in special education lessons are presented in Table 4. We found some clear discrepancies between teachers' judgements and the provision of special education services. In 11 cases, according to the teachers' scale-based judgements, the student belonged to both

TABLE 3 Correspondence between teachers' judgements and test results in assessing reading fluency

		Teachers' judgements			
Reading test results		Poorer reader	Average reader	Better reader	Total (N = 184)
Poorer reader	f_o	24	10	4	38
	f_e	8.47	14.87	14.66	38
	AR	6.80	-1.82	-3.99	
Average reader	f_o	14	43	25	82
	f_e	18.27	32.09	31.64	82
	AR	-1.52	3.32	-2.02	
Better reader	f_o	3	19	42	64
	f_e	14.26	25.04	24.70	64
	AR	-4.19	-1.92	5.50	
Total (N = 184)		41	72	71	184 ^a

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

^aIn the comparisons in Table 1, three students were removed because their audio recordings were defective.

TABLE 4 Relationships between the two teachers' assessment procedures and students' participation in special education lessons

First judgement		2nd judgement Identified as needing support			Special education lessons	
Scale-base judgement		In writing	In reading	Students (N = 184)	Participates	Does not participate
Poorer speller	Poorer reader					
Yes	Yes	+	+	18 ^b	17	1
		+	-	4	2	2 ^c
		-	+	1	0	1 ^c
		-	-	7 ^a	0	7
Yes	No	+	+	7	7	0
		+	-	3 ^b	3	0
		-	-	6	0	6
No	Yes	+	+	3	3	0
		-	+	1 ^b	1	0
		-	-	7	0	7
No	No	+	+	6 ^a	6	0
		+	-	4	3	1
		-	+	2	2	0
		-	-	115 ^b	0	115
Students (N = 184)		184			44	140

Abbreviations: “+”, need help; “-”, the need for help was not mentioned.

^aDisagreements between first and 2nd judgements in both reading and writing.

^bAgreements between first and second judgements in both reading and writing.

^cNon-participation was justified.

groups of poorer spellers and readers but did not receive a special educational service. Another 11 students participated in special education lessons focused on literacy development but were not identified as poorer readers or spellers based on scale judgements.

To go deeper, the relationships between the two teacher assessment procedures were analysed. The results of the first and second teacher assessment procedures were fully consistent in 137 cases (i.e., if, based on the scale judgement, the student belonged to the group of poorer spellers, poorer readers, or both, in the second assessment the teacher also indicated that the student needed additional assistance in spelling, reading, or both). We also found many partial discrepancies (i.e., the difference between the two assessment procedures was either reading-only or writing-only). In seven cases, based on the scale, teachers initially identified students as poorer in both reading and spelling, but in the second assessment noted that these students did not need additional reading or spelling assistance. None of these seven students participated in special education lessons. The opposite case occurred for six students: According to the scale judgements, these students did not belong in the poorer group for either spelling or reading, but teachers subsequently indicated that all six students needed additional help in both reading and writing. All these students also participated in special education lessons. With two exceptions, all students for whom the teacher indicated the need for additional support in the second assessment received special education services at school, even if they were not initially identified as students with poorer skills based on scale judgements, and vice versa.

3.3 | Teachers' and support specialists' judgements of skills of students participating in special education lessons

The final analysis focused only on students who received part-time special education services at school (SEN students) and whose literacy skills were evaluated by both teachers and support specialists. In the spelling part, the correlations between support specialists' judgements and test results as well as between teachers' judgements and test results were moderate: .711, $p < .01$, and .672, $p < .01$, respectively. Next, the consistency between the groups formed based on the spelling test results and the judgements was examined. Fisher's exact test indicated significant consistency between the categories of support specialist judgements and spelling test results ($p = .006$) as well as between the categories of teachers' judgements and spelling test results ($p < .001$). Regarding the support specialists'

TABLE 5 Correspondence between support specialists' judgements and test results in assessing SEN students' spelling skills

		Support specialists' judgements			
Spelling test results		Poorer speller	Average speller	Better speller	Total (N = 34)
Poorer speller	f_o	17	6	1	24
	f_e	12.71	9.18	2.12	24
	AR	3.2	-2.5	-1.5	
Average speller	f_o	1	4	1	6
	f_e	3.18	2.29	.53	6
	AR	-2.0	1.6	.7	
Better speller	f_o	0	3	1	4
	f_e	2.12	1.53	.35	4
	AR	-2.3	1.6	1.2	
Total (N = 34)		18	13	3	34

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

TABLE 6 Correspondence between teachers' judgements and test results in assessing SEN students' spelling skills

		Teachers' judgements			
Spelling test results		Poorer speller	Average speller	Better speller	Total (N = 34)
Poorer speller	f_o	22	2	0	24
	f_e	16.94	4.24	2.82	24
	AR	4.2	-2.2	-3.3	
Average speller	f_o	2	3	1	6
	f_e	4.24	1.06	0.71	6
	AR	-2.2	2.3	.4	
Better speller	f_o	0	1	3	4
	f_e	2.82	0.71	0.47	4
	AR	-3.3	.4	4.2	
Total (N = 34)		24	6	4	34

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

judgements, the AR examination indicated that the observed frequency was significantly higher than expected in only one category: poorer speller–poorer speller (Table 5). Regarding the teachers' judgements, the significant overlap between test results and teachers' judgements occurred in three categories: poorer speller–poorer speller, average speller–average speller, and better speller–better speller (Table 6). When judging the spelling skills of those students who belonged to the group of poorer spellers according to test results ($N = 24$), support specialists misclassified 7 (29%) students and teachers misclassified 2 (8%) students.

In the reading part, the correlations between judgements and test results were weaker than in the spelling part for support specialists (.415; $p < .01$) and teachers (.533; $p < .01$). Next, the consistency between the groups formed based on the reading test results and the judgements were examined. The consistency between the categories of support specialists' judgements and the test results did not contribute significant results (Fisher's exact test,

TABLE 7 Correspondence between support specialists' judgements and test results in assessing SEN students' reading fluency

		Support specialists' judgements			
Reading test results		Poorer reader	Average reader	Better reader	Total (N = 34)
Poorer reader	f_o	12	7	0	19
	f_e	10.06	8.38	0.56	19
	AR	1.3	-1.0	-1.1	
Average reader	f_o	5	7	0	12
	f_e	6.35	5.29	0.35	12
	AR	-1.0	1.2	-.7	
Better reader	f_o	1	1	1	3
	f_e	1.59	1.32	0.09	3
	AR	-.7	-.4	3.3	
Total (N = 34)		18	15	1	34

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

TABLE 8 Correspondence between teachers' judgements and test results in assessing SEN students' reading fluency

		Teachers' judgements			
Reading test results		Poorer reader	Average reader	Better reader	Total (N = 34)
Poorer reader	f_o	13	6	0	19
	f_e	10.06	8.38	0.56	19
	AR	2.0	-1.7	-1.1	
Average reader	f_o	4	8	0	12
	f_e	6.35	5.29	0.35	12
	AR	-1.7	2.0	-.7	
Better reader	f_o	1	1	1	3
	f_e	1.59	1.32	0.09	3
	AR	-.7	-.4	3.3	
Total (N = 34)		18	15	1	34

Abbreviations: AR, adjusted residuals; f_e , expected frequency; f_o , observed frequency.

$p = .112$). The consistency between the categories of teachers' judgements and test results was significant (Fisher's exact test, $p = .032$). Regarding support specialists' judgements, the AR examination indicated that the observed frequency was significantly higher than expected in only one category: better reader–better reader (Table 7). Regarding the teachers' judgements, the significant overlap between test results and teachers' judgements occurred in three categories: poorer reader–poorer reader, average reader–average reader, and better reader–better reader (Table 8). When judging those students who belonged to the group of poorer readers according to the test results ($N = 19$), support specialists misclassified 7 (37%) students and teachers misclassified 6 (32%) students.

4 | DISCUSSION

The aims of this study were to analyse the accuracy of teachers' and support specialists' judgements about students' literacy skills and to examine whether the provision of special educational services to a student affects teachers' judgements. The results indicated that teachers' judgements were somewhat more accurate for students' spelling skills than reading fluency. No statistically significant overestimation was found for low-skilled students, but the overestimation rate was still raised as an important issue as well as the fact that teachers' judgements of student skills could be influenced by the provision of special education services. Teachers' judgements were slightly more accurate than support specialists' judgements.

4.1 | Correspondence of teacher judgements and test results

First, the relationships between teachers' judgements and test results were analysed. According to our first hypothesis, we expected teachers' judgements to be more accurate for students' spelling skills than for their reading fluency. Similar to previous results (Freinberg & Shapiro, 2003; Hoge & Coladarci, 1989; Südamp et al., 2012), the correlation between judgements and test results was moderate, whereas the overall consistency between judgements and test results was somewhat higher in the spelling part. In particular, teachers classified low-skilled students more accurately in terms of spelling than in terms of fluency in reading. Thus, the first hypothesis was supported.

This finding can be explained by the fact that teachers receive information about spelling skills from students' written works, which are often done daily, and it is possible to analyse written works in more depth and preserve them for the purposes of assessing development. Regarding reading fluency, teachers can only receive information when they hear the student read aloud. Moreover, it is not known how often teachers ask third-grade students to read aloud and, thus, how much they purposefully collect information about the level of reading fluency. Previous research has highlighted concerns about the lack of purposefully collected data on students' skills by teachers (Vanlommel et al., 2017). This study did not examine how teachers monitor and evaluate students' skills on a daily basis, but inaccuracies in assessment may refer to teachers' intuition-based rather than data-based decisions.

Compared to the rate of poorer spellers identified by teachers (i.e., 83%), the level of identification of poorer readers was much lower (i.e., only 63%). This result is concerning because it can be assumed that teachers do not pay extra attention to supporting students who, in their opinion, have no difficulties related to reading fluency. According to the curriculum, Estonian students should be able to read fluently by the end of third grade (Government of the Estonian Republic, 2010), but most students are able to read rather fluently by the end of first grade (Soodla et al., 2015). Estonian's transparent orthography also facilitates learning to read (Soodla et al., 2003; van Daal & Wass, 2017). Because of these aspects, teachers may mistakenly believe that developing reading fluency does not require as much attention, which can lead to students who need help with reading fluency going unnoticed.

Our second expectation was that teachers tended to overestimate students with lower literacy skills. As demonstrated in several previous studies (Bates & Nettelbeck, 2001; Begeny et al., 2008; Feinberg & Shapiro, 2009; Soodla & Kikas, 2010), and as we also revealed, the number of students with lower skills is overestimated, but the rate of

overestimation was not statistically significant within our sample. Thus, Hypothesis 2 was not supported. However, our finding does not completely contradict the results of previous studies, as the misclassification of low-skilled students was proportionally still remarkable, particularly in the reading part, as previously mentioned.

One reason we did not find a statistically significant overestimation in the group of lower-skilled students may be that, in the last decade in Estonia, the education system has emphasized support for students with weaker skills and more attention is being paid to the implementation of the principles of inclusive education (Government of the Estonian Republic, 2010). In addition, more short- and long-term in-service training for teachers helps them identify and support learners with special needs. For example, the PISA results show fewer students have very low reading comprehension skills at the end of basic school in Estonia compared to other OECD countries (OECD, 2018), which may also indirectly indicate that teachers are able to notice and effectively support students having difficulties. Thus, the reason we did not find a systematic overestimation of low-skilled students may lie in the fact that, thanks to relevant in-service trainings and the emphasis on inclusion, teachers are more aware and notice students having difficulties rather well. In addition, most participating teachers had been teaching their students for three years. This longer teaching duration may also explain why, as previous studies indicated (Johansson et al., 2012), many teachers' judgements were quite—although not overly—accurate. Therefore, the results suggest that, in addition to routine practice, which is largely based on gathering information through daily observations (Begeny & Buchanan, 2010), teachers may need reliable assessment tools to adequately assess literacy skills, especially reading fluency. Currently, no such assessment tools are available for teachers in Estonia.

4.2 | Teachers' judgements and students' participation in special education lessons

We next analysed the relationships between teachers' judgements and the provision of special educational services. We assumed that teachers tended to underestimate the skills of students receiving special education services.

We analysed the relationships between the two teacher assessment procedures and found that the relationship was almost exclusively as follows: When special educational service was provided to a particular student at school, the teacher stated in the second assessment that the student needed additional help, even if she did not initially identify that student as a poorer speller or reader and vice versa. This tendency can be explained by the fact that, first, during the scale assessment, teachers were not directed to think about and indicate the students who receive special educational support at school. In the second assessment (i.e., when determining the need for help), they were asked to indicate those students currently receiving special educational support. Thus, the reminder that some students already receive special educational services might have influenced teachers when giving judgements—in the context of the present study, assessing the student as needing additional help. Therefore, Hypothesis 3 can be considered supported.

As several previous studies have found similar results (Campbell, 2015; Hurwitz et al., 2007; Soodla & Kikas, 2010), the main question is how aware teachers actually are about the difficulties with which the students struggle. When teachers' awareness is low, timely support may be delayed, and the consideration of the student's proximal development zone (Vygotsky, 1978) may be insufficient. It is particularly important that students with more severe learning difficulties, who in accordance with the principles of inclusive education spend most of their time in the classroom with other students, receive timely and appropriate support. Therefore, teachers' awareness of students' needs has grown over time and become crucial.

The results may also indicate that special education services may be provided to students who no longer need them (i.e., they have overcome their challenges). In the present study, this group may include students who continued to receive special educational services, but who, according to the teacher's initial judgement, did not belong to a poorer group. One explanation may be that, after identifying the initial problem, monitoring the student's further progress remains insufficient, thereby leading to students who really need support being deprived of it because the resource is being used inefficiently.

On the other hand, the provision of special educational services to students may be too labeling or even stigmatizing and can affect the way a teacher treats a student. For example, some studies have described how teachers believe that a student with learning difficulties will not be able to learn successfully with others in the classroom and there is nothing that can be done to support this student (Clark & Artiles, 2000; Woodcock & Vialle, 2010). Tournaki and Podell (2005) have also found that teachers' assessments and expectations tended to be higher for students whose reading skills were actually below grade level, but who were not formally identified as having learning difficulties and were not receiving special educational support. At the same time, teachers tended to underestimate skills of those students whose learning difficulties were identified and who were assigned special educational support. Due to this bias, teachers' expectations and demands of students may be low and even hinder student development.

Thus, the general suggestion is that teachers may also need information to prevent labeling as well as additional training in identifying, assessing, and supporting students with learning problems in an inclusive classroom. Moreover, if the teacher is not aware of the difficulties of the student receiving the special education service, it may indicate a lack of collaboration between the teacher and the support specialist, which has also been discussed in previous studies (Jurkowski & Müller, 2018; Khairuddin, Dally, & Foggett, 2016).

4.3 | Teachers' and support specialists' judgements about SEN students' skills

To study the third hypothesis, the relationships between teachers' and support specialists' judgements and test results related to SEN students' skills were analysed. We assumed that support specialists' judgements were more accurate than teachers' judgements based on an earlier study comparing the accuracy of teachers' and special educators' judgements that showed the latter's judgements were more accurate (Virinkoski et al., 2017). Contrary to the previous study and our expectations, we found that teachers' judgements were somewhat more accurate in assessing the level of both spelling and fluency in reading.

One possible reason for support specialists' more inaccurate judgements may be that, traditionally, support specialists work with students with learning difficulties in a separate room—mostly twice a week (Soodla et al., 2020); thus, they may not have an overview of the overall level of the class. As the general level of the class may influence the teacher's assessments (Kikas et al., 2018), it can be assumed that the same phenomenon could take place within a group of students attending special education classes. Support specialists might give higher judgements for those students whose skills are relatively good compared to the skills of other students within the small group. However, it can be assumed that those students' skill levels in the context of the whole class are still weak. Thus, for teachers, these students may be more clearly distinguished as having lower skills in the classroom context. Yet it must be emphasized that the judgements were collected on a scale containing short descriptions, which could have reduced the subjectivity. Furthermore, like teachers, support specialists do not have reliable assessment tools to assess age-appropriate levels of development.

Here, it is also important to highlight the collaboration between teachers and support specialists, which has often been found to be insufficient (Jurkowski & Müller, 2018; Khairuddin et al., 2016). According to the principles of inclusive education, a student exhibiting difficulties primarily studies with other students in the classroom. Therefore, support specialists should also spend more time in class observing the students, supporting them in the classroom, and consulting with teachers. As special educators estimate the potential of students with learning difficulties to cope in the classroom better than teachers do (Clark & Artiles, 2000; Vlachou et al., 2014), it can be assumed that collaborating in the classroom could be effective in engaging and supporting students. However, there is currently no common system in Estonia that regulates how the support specialists' work should be organized in schools, meaning it varies from school to school. For example, in some schools, special educators are more involved in counseling (for example, advising teachers and parents on supporting a student having difficulties) and the volume of direct teaching to students with learning difficulties is somewhat lower, whereas in other schools special educators mainly

teach and the volume of counseling is smaller. Thus, in one school, a special educator may support students with reading and writing difficulties more indirectly through the counseling of teachers and parents, but in another school, this may occur more directly (i.e., by teaching the student).

5 | LIMITATIONS AND CONCLUSIONS

Some limitations of the study should be considered. First, the sample size was relatively small and reduced the power of the analyses. Second, the unfamiliar researcher and test situation may have influenced students' test results, and therefore the skill level determined from the test results and the student's actual skill level may not overlap in some cases. Also, expert judgements made based on reading test performances were interpreted as reading test results and, thus, may be somewhat more subjective than spelling test results. We should also take into account that teachers and support specialists may have subjective interpretations of the terms used on the scales (e.g., "often makes mistakes" or "makes a lot of mistakes"). This in turn may have affected the results of comparing the accuracy of teachers' and support specialists' assessments. And finally, most participating teachers' classroom experience was relatively limited; their lack of experience may have affected their assessment skills. Despite these limitations, the study was conducted in a language setting with a transparent orthography, which is thus far less studied, thereby adding value through the findings. Other researchers from countries with a similar school system as in Estonia may also find the information useful.

The main conclusion from the study is that teachers and support specialists may need additional knowledge and skills for assessing students' literacy skills, particularly in reading fluency. In addition, collaboration between teachers and support specialists should be enhanced so that the principles of inclusive education can be applied more effectively and the labeling of students receiving special educational support can be reduced. The next step in this research field could be to observe lessons conducted in collaboration between teachers and support specialists and gather other data about collaboration, with the aim of finding effective practices to support students experiencing difficulties in a classroom context. Future research should also explore how teachers and support specialists make educational decisions about students—namely, how they collect data on students' skills, especially where reliable and standardized assessment tools are lacking.

ACKNOWLEDGMENT

This study was supported by EEA Financial Mechanism 2014-2021 Higher Education in Baltic Research Programme through the Project "Development and Enhancing the Teaching Quality of Inclusive Education Curriculum" (Contract No 36.1-3.4/289).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Maris Juhkam  <https://orcid.org/0000-0003-0363-1257>

Piret Soodla  <https://orcid.org/0000-0001-8413-5444>

Mikko Aro  <https://orcid.org/0000-0002-0545-0591>

REFERENCES

- 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&4), 149–178. <https://doi.org/10.1080/10627197.2006.965298>

- Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177–187. <https://doi.org/10.1080/01443410020043878>
- 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' judgements and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23, 43–55. <https://doi.org/10.1037/1045-3830.23.1.43>
- Begeny, J., & Buchanan, H. (2010). Teachers' judgments of students' early literacy skills measured by the Early Literacy Skills Assessment: Comparisons of teachers with and without assessment administration experience. *Psychology in the Schools*, 47, 859–868.
- Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy*, 44(3), 517–547.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In: M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255–296). New York: Macmillan.
- Clark, M. D., & Artiles, A. J. (2000). A cross-national study of teachers' attributional patterns. *The Journal of Special Education*, 34(2), 77–89. <https://doi.org/10.1177/002246690003400203>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Earl, L., & Louis, K. S. (2013). Data use: Where to from here? In K. Schildkamp, M. Lai, & L. Earl (Eds.), *Data-based decision making in education (studies in educational leadership)* (Vol. 17). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4816-3_11
- Eckert, T. L., Dunn, E. K., Coddng, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, 43, 247–265.
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgements in predicting oral reading fluency. *School Psychology Quarterly*, 18(1), 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 judgements 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>
- Government of the Estonian Republic. (2010). *Põhikooli ja gümnaasiumiseadus [Basic Schools and Upper Secondary Schools Act]*. Estonia: Riigi Teataja [State Gazette]. Retrieved September 16, 2020 from <https://www.riigiteataja.ee/akt/13332410>
- Government of the Estonian Republic. (2011). *Põhikooli riiklik õppekava [National Curriculum for basic schools]*. Estonia: Riigi Teataja [State Gazette]. Retrieved September 16, 2020, from <https://www.riigiteataja.ee/akt/129082014020>
- Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgements of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing: An Interdisciplinary Journal*, 15(7), 789–809. <https://doi.org/10.1023/A:1020985701556>
- Hint, M. (1998). Häälikutest sõnadeni [from speech sounds to words]. In *Eesti Keele Sihtasutus [Estonian Language Foundation]* (pp. 147–150).
- Hoge, R., & Coladarci, T. (1989). Teacher-based judgements of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.2307/1170184>
- Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgements of students' test performance. *School Psychology Quarterly*, 22(2), 115–144.
- Johansson, S., Myrberg, E., & Rosen, M. (2012). Teachers and tests: Assessing pupils' reading achievement in primary schools. *Educational Research and Evaluation*, 18(8), 693–711. <https://doi.org/10.1080/13803611.2012.718491>
- Jurkowski, S., & Müller, B. (2018). Co-teaching in inclusive classes: The development of multi-professional cooperation in teaching dyads. *Teaching and Teacher Education*, 75, 224–231. <https://doi.org/10.1016/j.tate.2018.06.017>
- Kallaste, L. (2016). Haridusliku erivajadusega õpilaste kaasava hariduskorralduse ja sellelega seotud meetmete tõhusus. Temaatiline raport: Statistiline ülevaade HEV levikust, kaasamisest ja tugimeetmete kasutamisest Eestis 2010–2014 [Effectiveness of inclusive education and related measures in Estonia. Thematic report: Statistical overview of the prevalence, involvement and use of support measures for students with SEN in Estonia 2010–2014]. <https://centar.ee/uus/wp-content/uploads/2017/01/Teemaraport-Statistiline-ulevaade-final.pdf>
- Karlep, K. (2000). Writing disabilities of Estonian children. *Trames*, 4(1), 53–78.
- Khairuddin, K. F., Dally, K., & Foggett, J. (2016). Collaboration between general and special education teachers in Malaysia. *Journal of Research in Special Educational Needs*, 16(1), 909–913. <https://doi.org/10.1111/1471-3802.12230>
- Kikas, E., Soodla, P., & Mägi, K. (2018). Teacher judgments of student reading and math 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>

- Kolne, K., Gonnerman, L., Marquis, A., Royle, P., & Rvachew, S. (2016). The factors contributing to teacher predictions of spelling ability, and the accuracy of their assessments. *Language and Literacy*, 18(1), 71–98. <https://doi.org/10.20360/G22P4H>
- 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>
- Lippus, P., Pajusalu, K., & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, 37(4), 388–396. <https://doi.org/10.1016/j.wocn.2009.07.002>
- OECD. (2018). *Program for International Student Assessment (PISA) results from PISA 2018 [PDF file]*. Paris: OECD Publishing. Retrieved September 20, 2020 from https://www.oecd.org/pisa/publications/PISA2018_CN_EST.pdf
- OECD. (2019). *TALIS 2018 results (volume I): Teachers and school leaders as lifelong learners*, TALIS. Paris: OECD Publishing. <https://doi.org/10.1787/1d0bc92a-en>
- Padrik, M., & Kikas, E. (2007). Special education in Estonia. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education, a reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals* (3rd ed., pp. 849–851). New York: Jhon Wiley & Sons.
- Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, 60, 425–433. <https://doi.org/10.1016/j.tate.2016.06.010>
- Ritchey, K. D., Coker, D. L., & Jackson, A. F. (2015). The relationship between early elementary teachers' instructional practices and theoretical orientations and students' growth in writing. *Reading and Writing: An Interdisciplinary Journal*, 28, 1333–1354. <https://doi.org/10.1007/s11145-015-9573-0>
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174.
- Shavelson, R., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgements, decisions, and behavior. *Review of Educational Research*, 51(4), 455–498. <https://doi.org/10.3102/00346543051004455>
- 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., Vija, M., & Pajusalu, R. (2013). Comparison of Estonian and Finnish word reading tests. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 9, 279–296. <https://doi.org/10.5128/ERYa9.18>
- 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. (2020). 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>
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Thiede, K. W., Brendefur, J. L., Carney, M. B., Champion, J., Turner, L., Stewart, R., & Osguthorpe, R. D. (2018). Improving the accuracy of teachers' judgements of student learning. *Teaching and Teacher Education*, 76, 106–115. <https://doi.org/10.1016/j.tate.2018.08.004>
- Tournaki, N., & Podell, D. M. (2005). The impact of student characteristics and teacher efficacy on teachers' predictions of student success. *Teaching and Teacher Education*, 12, 401–411. <https://doi.org/10.1016/j.tate.2005.01.003>
- van Daal, V. H. P., & Wass, M. (2017). First- and second-language learnability explained by orthographic depth and orthographic learning: A "natural" Scandinavian experiment. *Scientific Studies of Reading*, 1, 46–59. <https://doi.org/10.1080/10888438.2016.1251437>
- VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). The development and validation of a process for screening and referrals to special education. *School Psychology Review*, 32, 204–227. <https://doi.org/10.1080/02796015.2003.12086194>
- Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, 83, 75–83. <https://doi.org/10.1016/j.ijer.2017.02.013>
- Viise, N. M., Richards, H. C., & Pandis, M. (2011). Orthographic depth and spelling acquisition in Estonian and English: A comparison of two diverse alphabetic languages. *Scandinavian Journal of Educational Research*, 55(4), 425–453. <https://doi.org/10.1080/00313831.2011.587327>
- Virinkoski, R., Lerkkanen, M. K., Holopainen, L., Eklund, K., & Aro, M. (2017). Teachers' ability to identify children at early risk for reading difficulties in grade 1. *Early Childhood Education Journal*, 46(5), 497–509. <https://psycnet.apa.org/doi/10.1007/s10643-017-0883-5>
- 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>

- Vlachou, A., Eleftheriadou, D., & Metallidou, P. (2014). Do learning difficulties differentiate elementary teachers' attributional patterns for students' academic failure? A comparison between Greek regular and special education teachers. *European Journal of Special Needs Education*, 29(1), 1–15. <https://doi.org/10.1080/08856257.2013.830440>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Wheadon, C., Barmby, P., Christodoulou, D., & Henderson, B. (2020). A comparative judgement approach to the large-scale assessment of primary writing in England. *Assessment in Education: Principles, Policy & Practice*, 27(1), 46–64. <https://doi.org/10.1080/0969594X.2019.1700212>
- Woodcock, S., & Hitches, E. (2017). Potential or problem? An investigation of secondary school teachers' attributions of the educational outcomes of students with specific learning difficulties. *Annals of Dyslexia*, 67(3), 299–317. <https://doi.org/10.1007/s11881-017-0145-7>
- Woodcock, S., & Vialle, W. (2010). Attributional beliefs of students with learning disabilities. *The International Journal of Learning*, 17, 177–191. <https://doi.org/10.18848/1447-9494/CGP/v17i07/47160>

How to cite this article: Juhkam, M., Soodla, P., & Aro, M. (2022). How accurate are teachers and support specialists when judging students' literacy skills? Special educational service as an external factor influencing judgements. *Dyslexia*, 1–19. <https://doi.org/10.1002/dys.1725>