

Teacher Judgment of Early Reading Difficulty

Pledger Fedora

*Rose Institute for Learning and Literacy, School
of Education, Manhattanville College, Purchase,
New York, USA*

Abstract

Accurate teacher judgment of students' reading ability is critical for identification of students in need of intervention. Little research is available that examines the accuracy of teachers' judgments in identifying students at risk for early reading difficulty as students begin school. Findings of this study indicate that teachers are fairly accurate in their judgments, but the fact that reading ability provides the foundation for academic success suggests the need for a high degree of accuracy in their judgments of students' reading ability.

Key words: *at-risk students, elementary education, professional development, reading/emerging literacy, teacher education/certification.*

The ability to read and comprehend text provides the foundation for success in school. Research has established that for learning to read, the elementary grades are of primary importance (Adams, 1990). Longitudinal studies have indicated that more than 17% of students will experience reading problems in the elementary grades (National Institute of Child Health and Human Development [NICHD], 2000). Nationwide, a disproportionate number of poor readers are found in rural regions of America (Snow, Burns, & Griffin, 1998). Durham and Smith (2006) studied school readiness and found rural status to be associated with lower initial reading scores, particularly in certain ethnic minority groups and levels of socioeconomic status. This is likely the case due to the higher rate of poor children in rural areas, with minority children having double the rates of poverty

Portions of this article were adapted from P. M. Fedora (2014), "Single and double deficits in early readers in rural, low-wealth communities," Reading & Writing Quarterly: Overcoming Learning Difficulties, doi:10.1080/10573569.2013.866529, reprinted by permission of Taylor & Francis LLC, www.tandfonline.com.

Address correspondence to Pledger Fedora, Rose Institute for Learning and Literacy, School of Education, Manhattanville College, 2900 Purchase Street, Purchase, NY 10577, USA. E-mail: Pledger.Fedora@mville.edu

Teacher Judgment of Early Reading Difficulty

than nonminority children (O'Hare, 2009). According to Nagle, Hernandez, Emblar, McLaughlin, and Doh (2006), approximately 65% of all rural schools receive Title I funds. These figures indicate that many children attending rural schools may be at risk for early reading difficulty.

Mastery of early reading skills, such as phonological awareness (PA) and phonics, provides the foundation for successfully learning to read (NICHD, 2000). For students experiencing difficulty learning to read, early and intensive intervention is especially important for improving student outcomes (Fletcher, Lyon, Fuchs, & Barnes, 2007). In order to ensure equity in opportunity for students to participate successfully in schooling, teachers need to be well trained not only in reading development, the foundational skills of early reading, and evidence-based instructional models of reading, but also in methods for identifying students at risk. For students at risk for reading difficulty, early assessments are especially important (Invernizzi, Justice, Landrum, & Booker, 2004/2005) because research has demonstrated that once students fall behind, they tend to stay behind (Snow et al., 1998). Remediation is dependent on successfully determining which students are in need of intervention (Invernizzi et al., 2004/2005). To do this, teachers need the skills to accurately judge each student's reading ability and level of risk for early reading difficulty. Unfortunately, ensuring the development of these skills in teachers is especially challenging for schools in low-wealth, rural communities because of the lack of resources, difficulty attracting and retaining qualified teachers (Lee & Burkam, 2002; Skiba et al., 2008), and fewer opportunities for professional development in evidence-based reading instruction methods that might aid them in accurately assessing students' levels of risk for early reading difficulty. Due to these challenges, rural schools are less likely to have skilled teachers and resources to address severe reading difficulties than teachers in urban and suburban schools.

Teachers' judgments are important (Begeny, Krouse, Brown, & Mann, 2011; Meisinger, Bradley, Schwanenflugel, Kuhn, & Morris, 2009) because teachers make day-to-day educational decisions based on their judgments of students' abilities (Coladarci, 1986). Teachers' ability judgments can launch students into achievement trajectories early on, and these trajectories have a lasting influence (Alexander, Entwisle, & Dauber, 2003). Many factors influence teachers' judgments about students' abilities, including teachers' beliefs about education, the nature of the instructional task, and estimates of their students' cognitive abilities (Hoge & Coladarci, 1989); moreover, these judgments are highly correlated with teachers' instructional decision making (Hurwitz, Elliott, & Braden, 2007). Not only are teachers called on to make judgments about day-to-day occurrences (i.e., assigning reading groups), but they also are responsible for identifying students at risk for academic difficulties (Fuchs & Fuchs, 2006). Because of the importance of preventing reading disabilities, accurate teacher judgment of students' early reading ability is critical for identification of students in need of intervention.

Although there is a substantial body of research on teachers' judgments of student achievement, little research is available that focuses on teacher judgment of early reading skills. Recent research examining teachers' judgments of students' reading skills has indicated that teachers have difficulty judging students' reading levels (Begeny et al., 2011). Teachers tend to overestimate students' performance (Feinberg & Shapiro, 2009; Hamilton & Shinn,

Fedora

2003), especially for low-achieving students (Bates & Nettelbeck, 2001), and are less accurate in their judgments about these students (Madelaine & Wheldall, 2005). For example, Feinberg and Shapiro (2009) asked teachers to predict reading performance of average- and low-performing students using a rating scale and curriculum-based measures of reading fluency. A comparison of actual student scores with teacher-predicted scores indicated that teachers had lower levels of accuracy for lower achieving students. Begeny, Eckert, Montarello, and Storie (2008) investigated teachers' perceptions of first-, second-, and third-grade students' oral reading fluency and found that teachers generally were accurate when estimating the performance of students with strong oral reading fluency but less accurate when judging students with average to low oral reading fluency. In general, research on the accuracy of teachers' judgments in identifying students at risk (performing below grade level) for reading difficulty versus students not at risk (reading at or above grade level) has indicated that teachers' judgments tend to be accurate (Begeny et al., 2008). However, if teachers' judgments are not *reliably* accurate when identifying which children need early intervention, many children may not get the remediation they need as early as possible.

Of interest for this exploratory study was the accuracy of teachers' judgments of students' level of risk for early reading difficulty in first grade. For purposes of early identification and intervention, the accuracy of teachers' judgment of risk for reading difficulty may be more crucial than teachers' judgment of reading ability (Gijssel, Bosman, & Verhoeven, 2006). The immediate and long-term consequences of reading difficulty are significant, given the importance of learning to read and its role in academic achievement.

Early Reading Skills and Deficits

Research has demonstrated that the beginning stages of learning to read depend in great measure on PA (Adams, 1990; NICHD, 2000), rapid automatized naming (RAN), and oral language proficiency (Snow et al., 1998). Phonological-based difficulties have long been recognized as one of the primary sources of single-word decoding difficulties, which result in reading difficulties (Harm & Seidenberg, 1999). In recent years, research has established that fluent naming of visually presented stimuli is strongly associated with reading achievement (Clarke, Hulme, & Snowling, 2005). Further, comprehension is dependent on the ability to decode fluently and recognize single words automatically (Fletcher et al., 2007). Teachers need sufficient knowledge of reading development and the foundational skills of early reading (Moats, 2009) to recognize which students are in need of intervention.

To examine the reading skills of students, this study focused on the subprocesses of reading and employed the double-deficit hypothesis of Wolf and Bowers (1999). Those researchers posited that students struggling with learning to read may fall into one of three deficit subtypes: a PA deficit with average RAN ability, a RAN deficit with average PA ability, or a deficit in both PA and RAN abilities. Wolf and Bowers (1999) proposed that students who have a deficit in both PA and RAN are doubly impaired, so they may struggle more than peers with a single deficit (e.g., PA), and may be the most severely impaired. For purposes of analysis, students in the current study were sorted into one of two teacher-identified groups: nonstruggling or struggling. The struggling group was further sorted by the author into one of the three Wolf and Bowers (1999) deficit subtypes previously described.

Teacher Judgment of Early Reading Difficulty

The purpose of this exploratory study was to examine first-grade teachers' ability to identify students who were below grade level and not benefiting from classroom reading instruction versus students who were at or above grade level and benefiting from reading instruction. Specifically, this study sought to determine whether teachers would be able to accurately distinguish students who were at risk for early reading difficulty from those who were not. The following questions were addressed in this study: (a) Within the sample, are there single and double deficits in early reading skills of low-wealth, rural elementary students? (b) Will teachers be able to accurately differentiate students who are struggling with learning to read from those who are not struggling?

Method

A secondary analysis of the Targeted Reading Intervention (TRI) entry-level data was conducted for the present study. The TRI was a 2-year early literacy intervention study conducted in rural school districts in persistently poor counties with limited access to professional development opportunities. The districts and participating schools served economically and ethnically diverse communities predominately composed of African American and European American families. All schools were Title I schools with more than 75% of students eligible for free and reduced lunch. School size ranged from 357 to 402 students, and minority enrollment ranged from 50% to 75%. Schools were paired and matched based on demographic characteristics (e.g., number of students eligible for free and reduced lunch, school size, and minority enrollment). In Year 1, three schools from a single school district participated in the TRI study. In Year 2, two new schools from a second school district participated. Only entry-level data were included in the secondary analysis for the present study.

Participants

Twenty-two teachers participated in the TRI study. In Year 1, 10 first-grade teachers from three non-Reading First schools participated. In Year 2, 12 first-grade teachers from two Reading First schools participated. At each school, all first-grade teachers participated in the study. Teachers had a mean of 18 years of teaching experience with a mean age of 43; 65% were European American, 30% were African American, and 5% identified as "Other." Of the first-grade teachers, 10% held temporary teaching certification, 40% held regular teaching certification, and 50% held specific-grade certification. National Board Certification was held by 5% of the teachers.

Student participants included two cohorts of first-grade students ($N = 126$) from two school districts. In Year 1, the first cohort included 83 first-grade students, and in Year 2, the second cohort included 43 new first-grade students from the second school district. Of the 126 participating students, 52% were boys ($n = 66$) and 48% were girls ($n = 60$); 68% were ethnic-minority students ($n = 86$), and 32% were nonminority ($n = 40$). Students ranged in age from 5.8 to 7.4 years, with a mean age of 6.57 years.

Procedure for Selecting Student Participants

First-grade students were eligible to participate in the TRI study if they were not diagnosed with a disability and spoke English at home. Further eligibility to participate in the TRI study was determined by a two-step process. First, in September, teachers administered the state-mandated literacy assessment to each student in their classes. All literacy

Fedora

assessments were completed by the beginning of October in each year of the study. Second, using data from these assessments and the teachers' knowledge of their students' reading progress over the first 6 to 8 weeks of school, teachers completed a screening instrument for all students in their classes. Assessments were used to create a rank-ordered list of all students from lowest to highest performing in literacy. Next, teachers rated students by whether they were benefiting from regular classroom reading instruction, and whether they were below, at, or above grade level.

Students who the teachers identified as below grade level and not benefiting from classroom reading instruction were assigned to the struggling group. Students who the teachers identified as at or above grade level and benefiting from reading instruction were assigned to the nonstruggling group. Ten students per classroom—five students from the struggling group and five from the nonstruggling group—were randomly selected to participate in the original TRI study, which resulted in a stratified sample. Equal numbers of students from each of the identified deficit subgroups were selected to facilitate comparisons between the subgroups. For the purposes of the present study, students identified for the TRI study were further sorted by the researcher into one of Wolf and Bowers's (1999) deficit subgroups.

There were originally many more students eligible for the study than were included as participants. Lack of signed consent was the primary reason; only those with a signed parental consent form were able to participate. Other students were eliminated from the present study if they were age outliers, or if they were not able to complete all of the assessments. In these rural, low-wealth schools, many students were not able to complete the assessments.

Procedure

Before implementation of the intervention, all participating students in the first cohort ($n = 83$) were administered a battery of standardized assessments in the fall of the school year. Assessments were completed at each school by trained graduate students. Assessors had previous experience administering assessments and also participated in a 2-day training to practice administering assessments to students not participating in the study. In Year 2, the same selection and data collection procedures were followed for the second cohort ($n = 43$) from the two schools in a second rural county. For purposes of analysis for the present study, Cohort 1 and Cohort 2 were combined into a single cohort. Chi-square analyses were performed to determine the comparability of students in the two cohorts based on demographic characteristics.

Measures

Multiple measures were used to identify students exhibiting deficiencies and students who did not exhibit deficiencies. Assessment data from the Elision, Blending Words, and Sound Matching subtests of the Comprehensive Test of Phonological Processing (CTOPP) were used to measure PA (Wagner, Torgesen, & Rashotte, 1999). The Elision subtest examined students' ability to hear and manipulate sounds in spoken words. The students were presented with a word, asked to repeat the word, and then asked to say the word again, omitting a specific sound. Blending Words examined students' ability to integrate and say whole words after hearing parts (syllables or phonemes) of the words. Students were asked to blend phonemes into a word and say the word. The Sound Matching subtest examined

Teacher Judgment of Early Reading Difficulty

the students' ability to match sounds. The students were presented with a word and asked to identify a word with the same beginning or ending sound from a group of four orally presented words. The Elision, Blending Words, and Sound Matching subtests' scores were combined to generate the CTOPP Phonological Awareness Composite Score (PACS), which was used to measure each student's awareness of and access to the phonological structure of oral language. Wagner et al. (1999) reported test-retest reliability for Elision, Blending Words, Sound Matching, and PACS of .88, .88, .83, and .79, respectively, for students aged 5 to 7.

The Rapid Naming-Colors subtest of CTOPP (Wagner et al., 1999) was used to measure RAN. It measures retrieval of phonological information from long-term memory, as well as the ability to execute a sequence of operations quickly and repeatedly. The RAN task used a continuous naming paradigm in which students were presented with a printed page with randomly arrayed colors in rows or columns and asked to name the colors as quickly as possible while being timed. Higher scores on RAN tasks represent poorer RAN ability. Wagner et al. (1999) reported test-retest reliability of .78 for Rapid Naming-Colors for students aged 5 to 7.

The internal consistency reliability of the items on CTOPP subtests, except Rapid Naming-Colors, was calculated using Cronbach's (1951) coefficient alpha. Coefficient alphas for the Elision subtest were .95 for males and .94 for females. For Blending Words, the coefficient alphas were .92 for males and .91 for females. Coefficient alphas for Sound Matching were .94 for males and .93 for females. Alternate-form coefficients for the Rapid Naming-Colors subtest were .89 for males and .91 for females. All phonological tasks have demonstrated predictive validity with measures of reading (Torgesen & Wagner, 1998).

Assessment data from two subtests of the Woodcock-Johnson® III Diagnostic Reading Battery (WJ III DRB) were used to measure basic reading skills (Woodcock, Mather, & Schrank, 2004). The Letter-Word Identification (LWID) and Word Attack (WA) subtests of the WJ III DRB constitute the Basic Reading Skills (BRS) composite score and measure sight vocabulary, phonics, and structural analysis. The LWID subtest examines the ability of an individual to decode isolated words of varying difficulty. The first five LWID items involve symbolic learning, or the ability to match a rebus with a picture of the object. The remaining items measure reading identification skills in identifying isolated letters and words that appear in large type. The WA subtest examines the ability to decode phonetically regular nonsense words of varying difficulty. This subtest requires the individual to pronounce visually presented words in isolation. Test-retest reliability for LWID and WA for ages 5 to 18 was .94 and .91, respectively (Woodcock et al., 2004).

A standard score falling below the 25th percentile for national norms has been recommended by Siegel (2003) and used in other studies (e.g., McBride-Chang & Manis, 1996; Shany & Share, 2011) to categorize students into Wolf and Bowers (1999) deficit subtypes. Using this cutoff point, students were placed in one of the four subtypes for the current study: students with no deficit (ND), students with a deficit in RAN, students with a deficit in PA, and students with a double deficit (DD). For the PA group, the CTOPP PACS for the Elision, Blending Words, and Sound Matching subtests was used to determine deficit status. For the RAN group, the CTOPP Rapid Naming-Colors subtest score was used to determine deficit status. Students with deficits in both PA and RAN were moved to the DD group.

Table 1. Distribution of Students by Deficit Status

Deficit status	Number of students	%
ND	62	49.2
RAN	13	10.3
PA	26	20.6
DD	25	19.8

Note. ND = no deficit; RAN = rapid automatized naming; PA = phonological awareness; DD = double deficit.

Results

To determine whether patterns of deficits varied across students in the sample, students were divided into mutually exclusive subgroups based on the results of standardized assessments. Table 1 shows the distribution of the four subgroups within the sample.

Did teachers accurately distinguish students who were struggling with learning to read from those who were not struggling? It was expected that a statistically significant higher incidence of single and double deficits would be found among students whom teachers had previously identified as struggling than among those they had identified as nonstruggling. To determine whether a significant difference in the distribution of the deficit subtypes across group status existed, a two-way contingency table analysis was conducted. The two variables were deficit status with four levels (ND, RAN, PA, and DD) and group status with two levels (struggling and nonstruggling). Results indicated that deficit status and group status were significantly related [Pearson $\chi^2(3, N = 126) = 13.12, p = .004$, Cramér's $V = .32$]. This finding reflects the fact that when teachers identified a student as nonstruggling, 60% ($n = 41$) were true negatives (i.e., students identified as nonstruggling and who were not having difficulty learning to read); however, 40% ($n = 27$) were false negatives (i.e., students identified as nonstruggling but who were having difficulty learning to read). Of the 27 students who were false negatives, nine had a deficit in RAN, 11 had a deficit in PA, and seven had a DD. In contrast, when teachers identified a student as struggling, 64% ($n = 37$) were true positives (i.e., students identified as struggling and who were having difficulty learning to read); however, 36% ($n = 21$) were false positives (i.e., students identified as struggling who were not having difficulty). Risk status (struggling vs. nonstruggling) was significantly associated with subtype deficit. This would indicate that teachers were not significantly far off in their assessments of students' entry-level reading abilities. Table 2 shows the residuals

Table 2. Residuals for Group Status by Deficit Status

Deficit status	Nonstruggling			Struggling		
	Count	Expected	Residual	Count	Expected	Residual
ND	41	33.5	7.5	21	28.5	-7.5
RAN	9	7	2	4	6	-2
PA	11	14	-3	15	12	3
DD	7	13.5	-6.5	18	11.5	6.5

Note. ND = no deficit; RAN = rapid automatized naming; PA = phonological awareness; DD = double deficit.

Teacher Judgment of Early Reading Difficulty

for group status by deficit status. The largest residuals were found in the ND group and the DD group. More nonstruggling students were observed in the ND group than expected if group status and deficit status were unrelated. Similarly, more struggling students were observed in the DD group than expected. Therefore, students generally were classified by teachers as should be the case given their deficit status.

Discussion

When considering teacher judgment, it is important to remember that teachers were not attempting to place students in specific deficit subgroups when assigning them to the struggling and nonstruggling groups, as this was not the focus of the larger study from which these data were derived. As previously discussed, teachers completed a screening instrument in which they ranked all students in their class from lowest to highest performing in literacy. Low-performing students (struggling) were considered most at risk for early reading difficulty, and high-performing (nonstruggling) were considered least at risk.

Theoretically, if teachers were completely accurate in their judgments of their students' ability levels, then the ND group would be composed of only teacher-identified nonstruggling students, and the three deficit groups (RAN, PA, and DD) would be composed of only struggling students. Results of this study found the composition of all subgroups to be heterogeneous. Within the ND group, 66% were teacher-identified as nonstruggling; however, 34% were misidentified as struggling. Teachers misidentified 69% of the RAN group, 42% of the PA group, and 28% of the DD group as nonstruggling.

The high percentage of teacher-identified nonstruggling students in the RAN subtype may indicate that teachers did not recognize the potential problems associated with slow RAN skills and therefore did not view these students at risk for early reading difficulty. This mistake is easily made, especially when the student is an accurate but slow decoder (Wolf, 1999). This finding is particularly important in light of the impact that fluency has on comprehension (Pikulski & Chard, 2005) and academic achievement. Therefore, further research is needed in this area.

Additionally, teachers missed 42% of students with weak PA skills. This result may suggest that teachers in rural, low-wealth communities lack sufficient knowledge of the foundational skills of early reading, especially in the area of phonics (Moats, 2009), because teachers may not have learned to read using phonics and may have been inadequately prepared to teach PA and phonics in their teacher preparation programs (Lerner & Johns, 2009). As a result, they may have inadequate linguistic knowledge (Bos, Mather, Dickson, Podhajski, & Chard, 2001). To be able to analyze students' confusions and errors, teachers need to have a good understanding of the underlying structure of the English language (Moats, 2009; Washburn & Mulcahy, 2014). The teachers in the rural schools included in this study had a mean of 18 years of teaching experience. It is quite possible that the teachers may not have been trained in the most recent approaches to teaching reading. Lacking such preparation, teachers may misdiagnose or overidentify students at-risk for early reading difficulty. It is important to acknowledge that the challenges faced by teachers in rural, low-wealth communities are faced by all elementary teachers, but they may be exacerbated for teachers in rural, low-wealth schools due to decreased access to professional development opportunities.

Fedora

For high-achieving and average students, results of the present study were consistent with previous research that examined teachers' judgments (Begeny et al., 2008; Feinberg and Shapiro, 2003, 2009). However, unlike previous research that found teachers' judgments to be less accurate for low-performing students (Begeny et al., 2011; Demaray & Elliott, 1998), teachers in the present study were best at appropriately identifying students with the greatest need. This is important, given that the lowest achieving students would be those with the greatest need for intervention. Nonetheless, because of the importance of early identification and early intervention for all students in need of remediation, the percentage of students with deficits who were overlooked by the teachers is of concern. Of the students in the DD group, 28% were students whom the teachers identified as their better students—those they considered to be benefiting from classroom reading instruction and not struggling with learning to read. Because all of the students in the DD subtype group were students with a deficit in RAN and a deficit in PA, it is possible that teachers' lack of knowledge of the role of RAN skills may have implications for teachers' ability to recognize who is in the greatest need of assistance and who is not.

Limitations and Implications

This study contains several limitations that need to be discussed. First were those introduced by the method of sample selection. Due to the use of secondary data, it was not possible to control the method of selection. Cutoff scores used to identify students may have led to an overidentification of students in deficit groups, especially students in the PA group and the DD group. Additionally, parental consent was a condition of participation in the study; therefore, it is possible that students who were the neediest were also those students who were not able to participate due to lack of a signed consent form. Another limitation due to the method of sample selection was demographics. Minority students made up 68% of the sample; therefore, the students in this sample represent a group that is not generalizable to the larger population or even to other rural populations, due to the nature of the selection process and the distribution of students by minority status (Fedora, 2014a).

Second, there may be methodological consequences of sorting student participants into categories on the basis of correlated predictor variables (Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). Specifically, sorting students into groups based on correlated predictor variables may result in distortions in the distributions of mean scores for each subgroup and therefore affect interpretation of results. For example, if PA and RAN are positively correlated, then the group with a DD will have lower PA than the group identified as having just a single deficit in PA. As a result, the difference between the DD subtype group and the PA subtype group might be attributable to the DD group's more severe deficits in PA rather than their RAN deficit alone.

Despite these limitations, results from the present study have implications for researchers and teacher educators. The education and professional development of teachers responsible for preventing and remediating reading disabilities is a growing concern among teacher educators and researchers (Mather, Bos, & Babur, 2001; Moats, 2009; Moats & Foorman, 2003). Teachers often feel unprepared to address the instructional needs of their students, especially students in need of explicit instruction, because they may have limited understanding

Teacher Judgment of Early Reading Difficulty

of how students learn to read or why students experience difficulty with reading (Moats, 2009). If early identification and early intervention are to be maximally effective, then it is important for teacher education programs to align the content and methods of instruction with current research. Teacher education programs need to ensure that preservice teachers have the knowledge and skills to meet the diverse needs of students (Fedora, 2014b) by providing preparation not only in identifying students at risk, but also in the components of reading, structural analysis of English, and evidence-based instructional models of reading. Thorough preparation through course work or professional development in the use of assessments and interpretation of assessment results will help teachers recognize which students are struggling with early reading. This is a vital area because research has shown that preservice teachers do not always receive thorough preparation in the use of assessment instruments (Begeny & Martens, 2006). It is important to continue this line of research because of the pivotal role that teachers play in meeting the instructional needs of their students. Future research examining teacher education, continuing professional development, and their impact on teacher judgment should greatly expand our understanding of how teacher education mediates the accuracy of teacher judgment.

Conclusion

Low-wealth status and rural status converge to place students at considerable risk of early reading difficulty and the accompanying academic difficulty. In order to improve student outcomes, all teachers need to be very accurate in their judgment of students' early reading ability. To achieve this goal, teachers must have the resources and knowledge to accurately identify not only those students in need of extra help with PA, but also those students who are not fluent readers and, therefore, in need of extra help with fluency and automaticity. This can only be accomplished if teachers have the means to assess all areas of early reading ability and the knowledge of empirically based early reading instructional models that are designed to remediate an array of skill deficits.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary grades* (2nd ed.). New York, NY: Cambridge University Press.
- Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology, 21*(2), 177–187.
- 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. doi: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.
- Begeny, J. C., & Martens, B. K. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly, 21*(3), 262–285.

Fedora

- Bos, C., Mather, N., Dickson, S., Podhajski, B., & Chard, D. (2001). Perceptions and knowledge of preservice and inservice educators about early reading instruction. *Annals of Dyslexia, 51*(1), 97–120.
- Clarke, P., Hulme, C., & Snowling, M. (2005). Individual differences in RAN and reading: A response timing analysis. *Journal of Research in Reading, 28*(2), 73–86. doi:10.1111/j.1467-9817.2005.00255.x
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78*(2), 141–146.
- Demaray, M. K., & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13*(1), 8–24.
- Durham, R. E., & Smith, P. J. (2006). Nonmetropolitan status and kindergarteners' early literacy skills: Is there a rural disadvantage? *Rural Sociology, 71*(4), 625–661. doi:10.1526/003601106781262052
- Fedora, P. M. (2014a). Single and double deficits in early readers in rural, low-wealth communities. *Reading & Writing Quarterly: Overcoming Learning Difficulties*. Advance online publication. doi:10.1080/10573569.2013.866529
- Fedora, P. M. (2014b). What all reading teachers should know and be able to do. *Kappa Delta Pi Record, 50*(1), 24–30. doi:10.1080/00228958.2014.871687
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18*(1), 52–65.
- Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *Journal of Educational Research, 102*(6), 453–462.
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. New York, NY: Guilford Press.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99. doi:10.1598/RRQ.41.1.4
- Gijssel, M. A. R., Bosman, A. M. T., & Verhoeven, L. (2006). Kindergarten risk factors, cognitive factors, and teacher judgments as predictors of early reading in Dutch. *Journal of Learning Disabilities, 39*(6), 558–571. doi:10.1177/00222194060390060701
- Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*(2), 228–240.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491–528.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313.
- Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly, 22*(2), 115–144.
- Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004/2005). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research, 36*(4), 479–500. doi:10.1207/s15548430jlr3604_3

Teacher Judgment of Early Reading Difficulty

- Lee, V. E., & Burkam, D. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.
- Lerner, J. W., & Johns, B. H. (2009). *Learning disabilities and related mild disabilities: Characteristics, teaching strategies, and new directions* (11th ed.). Boston, MA: Houghton Mifflin Harcourt.
- Madelaine, A., & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with a curriculum-based measurement procedure. *International Journal of Disability, Development and Education*, 52(1), 33–42.
- Mather, N., Bos, C., & Babur, N. (2001). Perceptions and knowledge of preservice and inservice teachers about early literacy instruction. *Journal of Learning Disabilities*, 34(5), 472–482. doi:10.1177/002221940103400508
- McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing*, 8(4), 323–339. doi:10.1007/BF00395112
- Meisinger, E. B., Bradley, B. A., Schwanenflugel, P. J., Kuhn, M. R., & Morris, R. D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children. *School Psychology Quarterly*, 24(3), 147–150. doi:10.1037/a0017191
- Moats, L. (2009). Still wanted: Teachers with knowledge of language. *Journal of Learning Disabilities*, 42(5), 387–391. doi:10.1177/0022219409338735
- Moats, L. C., & Foorman, B. R. (2003). Measuring teachers' content knowledge of language and reading. *Annals of Dyslexia*, 53(1), 23–45. doi:10.1007/s11881-003-0003-7
- Nagle, K. M., Hernandez, G., Emblar, S., McLaughlin, M. J., & Doh, F. (2006). Characteristics of effective rural elementary schools for students with disabilities. *Rural Special Education Quarterly*, 25(3), 3–12.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
- O'Hare, W. P. (2009). *The forgotten fifth: Child poverty in rural America*. Durham, NH: Carsey Institute.
- Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58(6), 510–519. doi:10.1598/RT.58.6.2
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, 96(2), 265–282. doi:10.1037/0022-0663.96.2.265
- Shany, M., & Share, D. L. (2011). Subtypes of reading disability in a shallow orthography: A double dissociation between accuracy-disabled and rate-disabled readers of Hebrew. *Annals of Dyslexia*, 61(1), 64–84. doi:10.1007/s11881-010-0047-4
- Siegel, L. (2003). Basic cognitive processes and reading disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 158–181). New York, NY: Guilford Press.
- Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J., & Chung, C.-G. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children*, 74(3), 264–288.

Fedora

- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Torgesen, J. K., & Wagner, R. K. (1998). Alternative diagnostic approaches for specific developmental reading disabilities. *Learning Disabilities Research and Practice, 13*(4), 220–232.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing* [Examiner's manual]. Austin, TX: PRO-ED.
- Washburn, E. K., & Mulcahy, C. A. (2014). Expanding preservice teachers' knowledge of the English language: Recommendations for teacher educators. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 30*(4), 328–347.
- Wolf, M. (1999). What time may tell: Towards a new conceptualization of developmental dyslexia. *Annals of Dyslexia, 49*(1), 1–28. doi:10.1007/s11881-999-0017-x
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415–438.
- Woodcock, R. W., Mather, N., & Schrank, F. A. (2004). *Woodcock–Johnson III Diagnostic Reading Battery*. Itasca, IL: Riverside.