Research Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2022, Vol. 55(4) 292–305

DOI: 10.1177/00222194211036203
journaloflearningdisabilities.sagepub.com
SAGE

An Exploration of the Heterogeneous Nature of Reading Comprehension Development in First Grade: The Impact of Word and Meaning Skills

Emily J. Solari, PhD[1], Ryan P. Grimm, PhD[2], and Alyssa R. Henry, PhD[1]

Abstract

This exploratory study builds upon extant reading development studies by identifying discrete groups based on reading comprehension trajectories across first grade. The main goal of this study was to enhance the field's understanding of early reading comprehension development and its underlying subcomponent skills, with the intent of better understanding the development of comprehension in students who display risk for reading difficulties and disabilities. A sample of first-grade readers ($N = 314$) were assessed at three timepoints across the first-grade year. These data were utilized to derive empirical latent classes based on reading comprehension performance across the first-grade year. Reading subcomponent skill assessments (phonological awareness, word reading, decoding, linguistic comprehension, and reading fluency), measured in the fall of first grade, were compared across latent classes to examine how they related to growth across the first-grade year. Results suggest that there were four distinct latent classes with differential reading comprehension development, each of which could also be distinguished by the subskill assessments. These findings are presented within the context of the broader reading research base, and implications for practice are discussed.

Keywords

reading comprehension, at-risk readers, growth mixture modeling

Decades of research have shown that young children demonstrate varied reading abilities in the early primary grades and that these variations are predictive of later reading achievement (Francis et al., 1996; Nation et al., 2010; Roth et al., 2002; Storch & Whitehurst, 2002; Torgesen et al., 2001; Wagner et al., 1997; Wagner & Torgesen, 1987). Extant data also demonstrate that these variations in reading ability remain relatively stable over time (Skibbe et al., 2008), such that children who are struggling with reading early tend to struggle across elementary school and beyond, highlighting the importance of early identification and intervention for children who have impairments in reading skills. Converging evidence suggests that early identification and remediation of reading difficulties is essential for successful reading achievement across the life span. Research from the National Reading Panel (NICHD, 2000) demonstrates that waiting to identify and initiate supplemental reading instruction for struggling readers at the age of 9 years results in the majority of these struggling readers experiencing ongoing reading difficulties. Approximately, 75% of struggling readers who do not learn to read adequately in the early grades have persistent reading difficulties (Francis et al., 1996; Juel & Leavell, 1988; Torgesen & Burgess, 1998).

To adequately address the instructional needs of children who demonstrate difficulties in early reading, it is important to understand the heterogeneous nature of early reading abilities and specific skill deficits that may impact reading comprehension development and performance. Many studies have reported findings based on average reading trajectories (e.g., Francis et al., 1996; Nation et al., 2010; Storch & Whitehurst, 2002); however, it is estimated that between 5% and 10% of the student population do not perform at the average level and struggle with at least one aspect of reading development (Fletcher & Vaughn, 2009). Less is known about the reading development in this population of

[1]University of Virginia, Charlottesville, USA
[2]SRI International, Menlo Park, CA, USA

Corresponding Author:
Emily J. Solari, School of Education and Human Development, University of Virginia, Charlottesville, VA 22903-1738, USA.
Email: ejs9ea@virginia.edu

students, especially in the area of reading comprehension in the early primary grades. Most developmental reading studies have investigated early reading subcomponent skills based on whole sample means and have rarely investigated distinct subgroups of children within a sample. Broad reading studies that analyze sample means are informative to the field and inform the developmental science of reading; however, they do not provide a nuanced view of reading development that can inform the instructional needs of readers who are at risk for reading difficulties and disabilities. Importantly, very few studies have investigated the differences between subgroups of readers in their early reading comprehension performance and how subcomponent skills of reading comprehension may differ across the profiles of readers. This study investigates a broad range of subcomponent skills, including word level and meaning-related skills to add to the current literature base on early reading comprehension development. Investigations of this nature have important implications for early targeted reading intervention for students who are at risk for reading difficulties and disability.

Reading Comprehension and Underlying Component Skills

Ample studies have investigated reading comprehension and many have empirically demonstrated the importance of several early component skills and their essential contributions to its development. The Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990) postulates that reading comprehension is an interaction between two broad component skills: decoding and linguistic comprehension. The component skill of decoding refers to an individual's ability to automatically understand the written code of language to allow for successful word reading. In order for young students to decode accurately, they must first acquire the subcomponent skill of phonological awareness or an understanding that spoken and written words are made up of smaller elements of speech and that these sounds are represented by symbols in written text in a systematic way. Targeted intervention studies that center on enhancing word-level decoding skills have shown that when the gap between average readers and students at risk for reading difficulties is addressed aggressively in the early grades, later-developing reading problems may be prevented and, consequently, their severity reduced (Mathes & Denton, 2002; Vellutino et al., 2000). Converging evidence suggests that while decoding in essential for reading comprehension, its influence decreases across development (Garcia & Cain, 2014). Decoding is known to be an essential skill for adequate reading comprehension development; however, it is not sufficient as reading comprehension includes the complex integration of many other component and subcomponent skills (e.g., linguistic comprehension). Linguistic comprehension refers to an individual's ability to comprehend spoken language. This includes both lower order language skills such as single-word vocabulary and higher order skills such as syntax. The component skills (decoding and linguistic comprehension) and their subcomponent skills (e.g., phonological awareness and listening comprehension) begin to develop at an early age; for successful reading comprehension, both skills sets are required.

The Simple View was originally conceived to explain the reading comprehension development of students in later primary grades (third and fourth); however, various researchers have utilized it to investigate earlier reading development (e.g., Catts et al., 2003; Kendeou et al., 2009a; Oakhill & Bryant, 2003; Solari et al., 2018). Importantly, the original Simple View investigations (Hoover & Gough, 1990) contend that linguistic comprehension becomes more essential for reading comprehension performance as children reach middle elementary school; decoding skills are the most salient skill related to reading comprehension in the early elementary years, but their impact declines over time (Kendeou et al., 2009a; Kendeou, Savage, & van den Broek, 2009b; Kershaw & Schatschneider, 2012; Storch & Whitehurst, 2002). The framework has been used to describe variance in reading development in multiple age groups (Adlof et al., 2006, 2010; Kendeou et al., 2009a; Kershaw & Schatschneider, 2012; Nation et al., 2010; Tunmer & Chapman, 2012) and to describe subgroups of children who struggle with reading development (Catts et al., 2003, 2006; Grimm et al., 2018; Nation et al., 2010; Solari et al., 2018). Although not articulated in the Simple View, subsequent research has also demonstrated the unique contribution of reading fluency to the prediction of reading comprehension outcomes both in studies that investigate sample means (Adlof et al., 2006; Kim et al., 2010; Silverman et al., 2013; Tilstra et al., 2009) and those that specifically investigate at-risk samples (Chard et al., 2002; Eason et al., 2013).

Considerable research has been conducted on the development of early decoding and word reading skills during the early elementary years, including consequential intervention work that demonstrates that between 75% and 98% of first-grade students can acquire adequate word reading skills when they receive evidence-based reading instruction, including targeted supplemental reading interventions for the students who are most at risk (e.g., Mathes et al., 2005; McMaster et al., 2005). However, some intervention studies have shown that even when young students are successfully taught to decode words in the early grades, some of them fall behind in the later grades in reading comprehension (Catts et al., 2012; Compton et al., 2008; Slavin et al., 1996). Causal studies have demonstrated that, in addition to phonological abilities and decoding skills, reading comprehension is predicted by broader language assessments (Catts et al., 1999, 2001, 2006; Hogan et al.,

2014; Scarborough, 1990). Therefore, although word reading and decoding skills are essential components of successful reading comprehension, these word-level skills are not sufficient alone to comprehend written text. Yet it is unclear whether the Simple View of Reading framework functions differently across subgroups of early readers with varying skill levels. This study includes measures of phonological awareness, decoding, linguistic comprehension, and oral reading fluency to identify significant differences among subgroups of first graders who differ in terms of early reading comprehension development. Some studies have suggested that many students demonstrate risk for both word reading and reading comprehension–related skills (Catts et al., 2003; Grimm et al., 2018; Solari et al., 2018). However, considerably less attention has been paid to early linguistic comprehension difficulties. The attention paid to early decoding skills is important but may not be singularly responsible for early reading comprehension difficulties. To better understand the full picture of early reading risk in young elementary-age children, we must have a greater awareness of how decoding and linguistic comprehension develop across subgroups of readers. It is important to have a solid understanding of how these skills begin to develop in the early elementary years to develop adequate and efficient early reading interventions.

Use of Growth Mixture Modeling (GMM) in the Context of This Study

Although longitudinal studies of reading comprehension development and the influences of its subcomponent skills are not uncommon in the literature, there are few studies that utilized GMM as an analytic approach (e.g., Boscardin et al., 2008). GMM provides specific advantages compared with techniques that analyze sample means. First, subgroups within a given sample are empirically identified and the data are used to differentiate participants instead of researcher-imposed cutoff scores. Francis et al. (2005) argued such cutoff scores could lead to instability as students who have scores that hover near the cutoff boundary may switch subgroups. Conversely, the data-driven approach used in GMM classifies students into relatively homogeneous subgroups based on similar patterns of development over time. If conceptually meaningful, such subgroups can be reflective of important differences within a sample that may be obfuscated when analyses are restricted to sample means. Furthermore, while reading comprehension development can be modeled as a continuum, identifying and examining subgroups can provide teachers with a practical heuristic that aligns well with teaching practices, especially in early elementary years. It is not uncommon for early elementary teachers to differentiate reading instruction by dividing their classrooms into groups of students with similar skill profiles. Therefore, empirically identifying subgroups may help teachers target instruction more efficiently and accurately.

A second advantage of GMM concerns specific aspects of the modeling process. GMM is a model-based approach, meaning that models can be replicated with independent samples and results can be compared to confirm or reject results. Related, model fit statistics are provided, which allows multiple competing GMMs to be analyzed to identify the preferred model. In addition, as with other analyses such as latent growth curve modeling, auxiliary variables can be included in GMMs, but these effects are examined within each subgroup (and comparisons can be made across subgroups) rather than the sample as a whole. This provides more nuanced information with respect to the effects of the auxiliary variable. For example, these variables might demonstrate significant effects for one subgroup, but not others. Such a nuanced understanding would help teachers tailor instruction and allocate resources efficiently.

Current Study

Early identification of subgroups of at-risk readers can be challenging, especially when considering variables beyond early word reading. Furthermore, the majority of previous work has relied upon arbitrary decisions of cutoff scores groups instead of utilizing empirically based procedures. This study expands on this literature in two ways. First, we use statistical techniques to empirically identify categorically distinct developmental trajectories of reading comprehension in first-grade students. Second, we compare performance between the latent classes on subcomponent skills of reading to determine which of these skills significantly differ across latent classes. Whereas prior research has examined early predictors of reading comprehension, a unique aspect of this study is that it investigates whether these predictors have differential impacts on reading comprehension development, depending on subgroup status. That is, although the students in this sample are of similar ages, the predictive capacity of reading subskills may vary, depending on a student's level of reading comprehension. While the trajectories of typically developing students would likely show average or above average scores in reading comprehension throughout the year, we expect that subgroups of readers who are struggling with reading comprehension at the beginning of the year will continue to struggle as they may not be receiving the targeted instruction in listening comprehension skills needed to boost overall reading comprehension achievement. This study provides a more nuanced understanding of the roles of reading comprehension subskills because it explicitly accounts for sample heterogeneity while modeling the relationships between the predictors and reading comprehension

development. To this end, this study asks the following research questions:

Research Question 1 (RQ1): How many categorically distinct reading profiles emerged from this sample of first graders?
Research Question 2 (RQ2): Were there significant differences on reading comprehension between the distinct profiles at each of the three timepoints?
Research Question 3 (RQ3): Did the reading profiles differ in terms of early subcomponent skills of reading (phonological processing, word reading, decoding, linguistic comp, and reading fluency)?

Method

Setting and Participants

The sample for this study comes from a larger randomized control trial in which the treatment group received intensive reading intervention. For the purposes of this article, included in these analyses is the untreated randomly assigned control sample; therefore, children in these analyses did not receive specialized reading intervention but only the reading instruction that the school normally provided. Children came from 64 first-grade classrooms located in two states, California and Texas. The 64 classrooms were situated in 25 schools; all classrooms provided all instruction in English only. The student sample ($N = 300$) was randomly selected from the participating classrooms; parent consent was obtained in accordance with university institutional review board (IRB) procedures. Fifty-six percent of the sample were female and 61.6% qualified for free or reduced-price lunch status. Forty-two percent identified as Latinx, 27.6% as Black or African American, 14.6% as White, 4.7% as Asian, and 9.6% as mixed or Other, with 1.7% missing. Finally, 18.3% of the sample was classified as English language learners by their schools.

Assessment Procedures

Assessments occurred at three timepoints across their first-grade school year, once in the fall (September), at mid-year (January), and at end of year (May). The second author trained the data collectors; initial training took place across 2 days and provided an overview of the assessments, guidance on giving the assessments, and practice with peers and trainers. Assessors further practiced administering assessments and participated in a "check-out" session with the second author in which a scripted check-out protocol was utilized to determine whether each assessor was prepared to work directly with students. All assessors met a minimum threshold of 90% accuracy on each assessment. In addition, the second author double-checked all assessments in the field during testing sessions and corrected any mistakes before the students returned to class.

Measures

Phonological awareness. This study used two subtests from the Comprehensive Test of Phonological Processing–Second Edition (CTOPP-2; Wagner et al., 2013) to assess phonological awareness: Elision and Blending. The elision subtest requires individuals to repeat a verbally presented word while omitting a specified sound (e.g., "say bold. Now say bold without /b/"). Blending requires students to combine separate sounds into a word. For example, the student may be presented with, "What word do these sounds make: t- oi?" The correct response would be toy. The examiner's manual reports a test–retest reliability of .88 for both subtests with this age group.

Word reading. This study assessed word reading and decoding skills using the Letter Word Identification (LWID) and Word Attack (WA) subtests of the Woodcock Johnson–IV (WJ-IV; Mather & Jaffe, 2016). The LWID subtest assesses the ability to identify letters and read words presented in a list format. The WA subtest measures phonological decoding through pseudoword reading. Both subtests are untimed; administration ends when the student misses six responses. Reliability coefficients for the LWID for the age range in this study are between .96 and .98; for the WA they are .94 and .96.

This study also administered the Test for Word Reading Efficiency–Second Edition (TOWRE-2; Torgesen et al., 2012), which consists of Sight Word Efficiency (SWE) and Phonemic Decoding Efficiency (PDE). For each subtest, individuals are presented with a list of real words (SWE) or pseudowords (PDE) that get progressively more difficult and are asked to read as many words as they can in 45 s. Their score, for each subtest, is the total correct in the allotted time.

Fluency. We administered two measures of fluency. First, we used the Gray Oral Reading Test–Fifth Edition (GORT-5; (Wiederholt & Bryant, 2012) rate subscale to measure fluency of reading connected text. The rate score is normed based on the number of seconds it took the student to read the passage. The examiner's manual reports a Cronbach's alpha value of .86.

Second, we administered the oral reading fluency subtest of Dynamic Indicators of Basic Early Literacy Skills (DIBELS), which is an individually administered test of fluent reading in connected text. Students' fluency is evaluated on 1-min timed reading samples. The number of correct words per minute for each passage is recorded. Errors are defined as mispronunciations, omissions, substitutions, reversals, and hesitations longer than 3 s. At each timepoint,

students read two passages, yielding a total of six different passages across first grade. We calculated all possible permutations of the presentation order of the passages and randomly assigned permutations to students; no student in this study received the same order of passages as any other student. Reliability coefficients ranged from .66 to .97.

Linguistic comprehension. We used two measures of listening comprehension. First, we used the Understanding Spoken Paragraphs (USP) subtest from the Clinical Evaluation of Language Fundamentals (CELF; Semel-Mintz et al., 2003). This subtest assesses a student's ability to interpret factual and inferential information presented orally. The test manual reports a reliability using Cronbach's alpha of .69 and .65, and a reliability using the split-half method of .74 and .73, for 6-year-olds. Second, we used the Qualitative Reading Inventory–Fifth Edition (QRI-5; Leslie & Caldwell, 2011). Assessors read a passage to students and asked them six comprehension questions regarding implicit and explicit details. We used the total number of comprehension questions answered correctly as this assessment is non-normed. Cronbach's alpha was calculated to be .65 for the study sample.

Reading comprehension. This study used the GORT-5 (Gcomp) reading comprehension subscale to assess reading comprehension. This assessment comprises 16 progressively more difficult reading passages. The passages are read aloud by a child and are followed by five open-ended comprehension questions. The test was discontinued when a student was deemed unable to read a passage fluently based on the test's stopping criterion. The technical manual reports Cronbach's alpha of .92 for the age range in the study.

Data Analysis Plan

We used GMM (Muthén & Shedden, 1999) to empirically identify discernible latent classes of longitudinal reading comprehension development from fall through spring of first grade. Mplus syntax for the final model in this study is available in online supplemental Figures S1 and S2. GMM is an iterative technique and this study began by fitting a model with one latent class, and then increasing the number of latent classes by one in subsequent models. Latent classes are identified by heterogeneity in the intercept (i.e., the average GORT-5 comprehension score at the first timepoint) and slope parameters (i.e., shape of the developmental trajectory). That is, each latent class has its own estimates of average intercept and slope parameters. GMM assigns individuals to the latent class for which they have the highest posterior probability of membership. As in other mixture models, the latent classes were considered mutually exclusive and exhaustive. All models were conducted in Mplus 8.1, using full information maximum likelihood estimation with robust standard errors. This estimator included students in the GMM as long as they had a score on the GORT reading comprehension measure for at least one timepoint. For the auxiliary variable analyses (see the following), students were excluded if they were missing scores only on the particular variable being analyzed. The greatest amount of missing data occurred with GORT-5 reading comprehension at Timepoint 3, with 5.1% of the full sample missing scores on that measure (see Table 1). This study clustered students at the classroom level to account for nesting.

In addition to varying the numbers of latent classes, this study also examined two other types of model specifications. First, this study examined different developmental trajectories to best depict the shape of reading comprehension development for the full sample; both linear and latent basis (Raykov & Marcoulides, 2006) models were conducted. Second, as GORT reading comprehension was a continuous variable, this study examined different variance–covariance structures as recommended by Masyn (2013). Masyn refers to the most restrictive structure as "class-invariant, diagonal," which constrains the within-class indicator covariances to 0 and holds the variances of each indicator to be equal across latent classes. This is the default specification in Mplus. The second structure, "class-varying, diagonal," also fixes the within-class indicator covariances to 0, but allows the indicator variances to be freely estimated and differ across latent classes. The "class-invariant, unrestricted" structure allows both the indicator covariances and indicator variances to be estimated, but they are constrained to equality across latent classes. Finally, the least restrictive structure, "class-varying, unrestricted," freely estimates indicator covariances and variances, and allows them to differ across the latent classes. Utilizing different structures has the potential to alter the number and shapes of the latent classes, which can affect their substantive interpretation and inform considerations regarding parsimony of the competing models. We refer the reader to Masyn (2013) for additional details. We conducted a set of six GMM latent classes, using each of the developmental trajectory specifications and Masyn's variance–covariance specifications. Therefore, five latent classes by two developmental trajectory specifications by four variance–covariance structures yielded 48 models that were compared. However, for brevity, this study only presents the fit statistics of the set of models utilizing the linear developmental trajectories and the *diagonal, class-invariant* specifications as these were chosen as the final set of specifications.

Among all models, fit statistics, as well as substantive interpretation, for each model were compared to identify the preferred unconditional model (i.e., without auxiliary variables). The Bayesian information criterion (BIC;

Table 1. Descriptive Statistics of GORT-5 at All Timepoints and Subcomponent Skills From the First Timepoint.

Variable	*n*	*M*	*SD*	Minimum	Maximum	Skewness	Kurtosis
GORT Timepoint 1	300	5.39	6.21	0	24	1.00	−0.02
GORT Timepoint 2	290	7.67	6.55	0	24	0.48	−0.92
GORT Timepoint 3	284	10.88	7.56	0	29	0.08	−1.03
WJ letter word ID	299	88.90	18.04	40	138	−0.02	0.31
WJ word attack	299	95.48	19.05	40	132	−0.77	0.42
TOWRE SWE	299	90.39	18.14	55	142	0.75	0.37
TOWRE PDE	298	88.70	16.52	55	145	0.64	−0.05
GORT fluency	299	7.11	2.81	1	15	0.67	−0.30
DIBELS ORF	300	121.50	25.25	0	121.5	1.68	2.27
CTOPP phonological processing	298	96.44	16.13	55	145	0.20	0.40
CELF USP	298	8.18	3.69	1	15	−0.38	−0.77
QRI listening comprehension	300	2.92	1.62	0	6	−0.12	−0.77

Note. GORT scores are unstandardized to capture growth. All other measures are standard scores except DIBELS and QRI, which are non-normed measures. GORT-5 = Gray Oral Reading Test–Fifth Edition; WJ = Woodcock Johnson; TOWRE = Test for Word Reading Efficiency; SWE = sight word efficiency; PDE = phonemic decoding efficiency; DIBELS ORF = Dynamic Indicators of Basic Early Literacy Skills–Oral Reading Fluency; CTOPP phonological processing = phonological processing composite; CELF USP = Clinical Evaluation of Language Fundamentals–Understanding Spoken Paragraphs subtest; QRI = Qualitative Reading Inventory–5.

Schwartz, 1978), which is currently the most trusted fit index of mixture models (Nylund et al., 2007) and the adjusted BIC (ABIC) are both interpreted as the minimum value indicating the preferred model. This study also utilized the Lo–Mendell–Rubin (LMR) test to examine whether the addition of another latent class resulted in a significant improvement in model fit. That is, the LMR compares a model with k classes and a model with $k - 1$ classes. The model with $k - 1$ classes is preferred if the p value for the model with k classes is nonsignificant (i.e., the model with one additional class did not significantly improve the fit of the model). We also examined the Bayes Factor (BF; Masyn, 2013) and correct model probability (cmP; Maysn, 2013). The BF compares a model with k classes with a model with $k - 1$ classes and ranges of values provide degrees of the strength of evidence for the $k - 1$ class model; values between 1 and 3 are weak evidence, values between 3 and 10 are moderate evidence, and values greater than 10 are strong evidence. The cmP provides the probability of each model being preferred against the full set of models under consideration, with the highest probability indicating the preferred model. While not a fit statistic, this study also examined the entropy values for all models, which summarizes how well individuals are assigned to latent classes. Entropy values range from 0 to 1, and values above .80 are considered indicative of strong classification (Clark & Muthén, 2009).

After we identified the preferred model, we added auxiliary variables using the BCH approach (Asparouhov & Muthén, 2014; Bakk et al., 2013; Bolck et al., 2004), so they would not influence class enumeration (Bakk & Vermunt, 2016). The auxiliary variables were phonological awareness, word reading, decoding, linguistic comprehension, and reading fluency measured at the first timepoint. This study included these variables to identify skills in the beginning of first grade, which were related to overall development of reading comprehension during the entirety of the school year. This approach applies weights to each individual based on their posterior probabilities of latent class membership, which reduces bias when estimating class-specific means of auxiliary variables. Once the auxiliary variables were included, the BCH approach estimated and compared class-specific means as a multiple group analysis, with the latent classes treated as observed subgroups. We examined all pairwise comparisons for significant differences across auxiliary variables, using a series of Wald tests provided by the BCH approach. We used standard scores for each variable except DIBELS and QRI that are non-normed; thus, raw scores were used for those two variables.

Results

Descriptive Statistics

Descriptive statistics for the GORT-5 reading comprehension measure from all three timepoints, as well as the covariate results from the first timepoint, are presented in Table 1. GORT-5 reading comprehension scores are non-normed to capture growth, whereas covariate scores are normed (except DIBELS and QRI that are non-normed measures) to describe the sample relative to the population.

GORT-5 reading comprehension scores increased approximately 3 points between each timepoint, suggesting linear growth. The mean scores for the normed variables were within the average range. However, analysis of mean

Table 2. Fit Statistics of the GMMs With a Linear Growth Trajectory and the Class-Invariant, Diagonal Structure.

No. of profiles	No. of param	LL	BIC	ABIC	LMR *p* value	BF	cmP
1	6	−2745.70	5525.63	5506.60	N/A	<.01	<.01
2	9	−2567.46	5186.26	5157.72	<.001	<.01	<.01
3	12	−2490.69	5049.82	5011.76	<.001	<.01	<.01
4	15	−2452.48	4990.52	4942.95	**.002**	<.01	<.01
5	18	−2422.81	4948.28	4891.19	.185	<.01	<.01
6	21	−2393.95	4905.67	4839.07	.082	N/A	.99

Note. The value in boldface represents a preferred model for a given fit index. GMM = growth mixture modeling; LL = log-likelihood; BIC = Bayesian information criterion; ABIC = adjusted BIC; LMR = Lo–Mendell–Rubin Likelihood Ratio Test; BF = Bayes factor; cmP = correct model probability.

scores may obfuscate struggles experienced by students who deviate from the average; thus, this study utilized GMM to empirically identify these students.

Identifying the Number of Latent Classes

Linear models of growth consistently provided good fit to the data and this was consistent with the change in mean GORT-5 scores over time. The latent basis models also demonstrated near-linear growth, although the class-specific mean GORT-5 scores at the middle timepoint were freely estimated instead of fixed (as in the linear models). Moreover, the fit statistics for linear and latent basis models with corresponding numbers of latent classes were nearly or exactly identical depending on the number of classes. Thus, due to similar fit statistics and parsimony considerations, the linear specification was chosen. Next, using the linear specification, four sets of models using different variance–covariance structures were conducted. Two of the four structures, which Masyn (2013) refers to as "class-varying, unrestricted" and "class-varying, diagonal" experienced convergence problems and were removed from consideration. The model using the "class-invariant, unrestricted" structure had errors with models consisting of three or more classes. Thus, this study used the "class-invariant, diagonal" structure, which fixes the covariance between the intercept and slope to 0 and constrains their variances to equality across latent classes. When comparing GMMs utilizing the linear and latent basis specifications of the growth trajectory, the linear models consistently showed better fit than the latent basis models. The final set of specifications was narrowed down to linear GMMs with class-invariant, diagonal variance–covariance structures. Finally, the intercept variances were negative across latent classes (as they were constrained to equality across classes), but the estimates were nonsignificant, so they were fixed at 0 in the final models.

Fit statistics for the GMMs with one to six latent classes using these specifications are presented in Table 2. The LMR supported the four-class model. In addition, there was no overlap among classes and they appeared to be well-separated. The entropy value for this model was .94, which is considered high.

As only one of the fit statistics provided clear evidence of a model, this study also examined each model for substantive considerations to ensure that the latent classes were conceptually viable and not redundant. The three-class model consisted of a class performing well below average across all three timepoints, a class performing in the average range, and a class performing well above average. The four-class model provided a more nuanced picture of struggling students as there were two classes performing below average. Moreover, the fourth class was further validated by the covariate results (see the following) as there were significant differences between all four classes. Given that the four-class model had some statistical support from the fit indexes and enjoyed substantive validation, it was chosen as the preferred unconditional model.

Figure 1 provides a visual depiction of the four-class model. To label the latent classes, the class-specific mean GORT-5 comprehension raw scores at each timepoint were matched with corresponding scale scores. This allowed the classes to be labeled in reference to test norms. The class at the top of the plot consistently performed above average throughout the year although they showed no growth. This class was called *Above Average* and consisted of 5.1% of the sample. The class beneath it, with diamond markers, consisted of 17.9% of the sample and was labeled *Average*. The third class from the top, with square markers, was termed *Low Average* and made up 25.7% of the sample. Finally, the class at the bottom was labeled *Low* and was the largest class with 51.3% of the sample. The class-specific GORT means can be seen in Table 3.

Relating Latent Classes to Subcomponent Reading Skills

Multiple auxiliary variables of phonological awareness, word reading, reading fluency, and linguistic comprehension were subsequently added to the GMM to identify which variables from the beginning of first grade would be significantly different between pairs of latent classes and

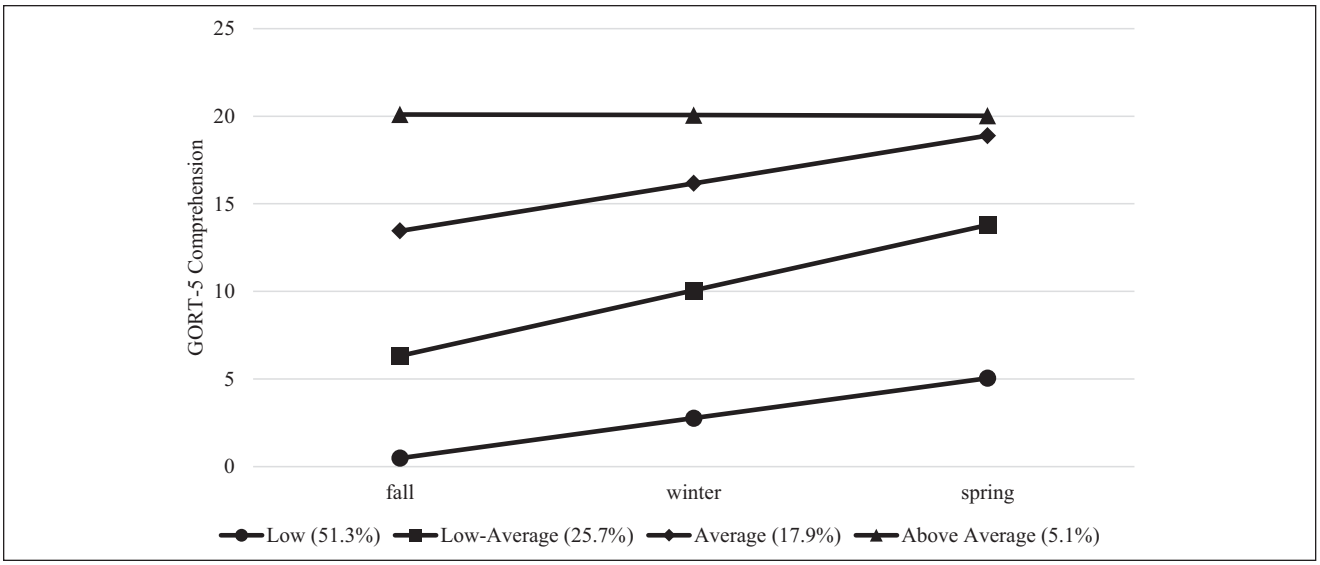

Figure 1. Plot of the final four-class model depicting heterogeneous growth patterns of GORT-5 reading comprehension across fall, winter, and spring of first grade.
Note. GORT-5 = Gray Oral Reading Test–Fifth Edition.

Table 3. Class-Specific Means of GORT Comprehension at Each Timepoint.

Latent class	Low	Low-average	Average	Above average
GORT Timepoint 1	0.49	6.33	13.45	20.10
GORT Timepoint 2	2.77	10.06	16.17	20.06
GORT Timepoint 3	5.05	13.79	18.89	20.02

Note. GORT = Gray Oral Reading Test.
All pairwise comparisons at all timepoints were significantly different at $p < .001$.

thus provide information as to which skills would be related to differences among students. This was accomplished using the BCH approach described above, which provided class-specific means for each variable and tests of statistically significant differences for all pairwise comparisons of latent classes. Results are presented in Table 4.

All pairwise comparisons for all auxiliary variables were statistically significant except for three instances and these three instances all occurred when comparing the *Average* and *Above Average* classes. The *Average* and *Above Average* classes were not significantly different on PDE, phonological processing, and QRI. In all other cases, the class-specific means were ordered as would be intuitively expected.

The *Low* class had the lowest means for all auxiliary variables and all standard scores were below the average range except phonological processing, but this was nearly below average. The *Low Average* class had significantly greater scores than the *Low* class, but remained below the *Average* and *Above Average* classes. The *Low Average* class scored within the average range on standardized variables, but, except for WA, all standard scores were below the mean score (i.e., 10 or 100 depending on the assessment). The *Average* class scored above the mean score on standardized variables, but within the average range. Finally, the *Above Average* class scored approximately 1 *SD* or greater than the mean score on all standardized assessments, except PDE and phonological processing, which were both on the high end of the average range. As these results were consistent with practical expectations and aligned well with the rank order of the latent classes, these findings were interpreted as providing conceptual validation to the latent classes.

Discussion

The main goal of this article was to investigate trajectories of reading comprehension development across the first-grade year; the data in this study were analyzed to enable a more nuanced understanding of the heterogeneous nature of reading comprehension development during first grade. The sample means and standard deviations indicated average range performance across all reading measures; however, results of the GMM indicated that four distinct trajectories of reading comprehension development were

Table 4. Differences Between All Pairwise Comparisons of Subcomponent Reading Skills.

			Mean differences (row minus column)		
Growth class	Subcomponent skills	*M*	Low	Low-average	Average
Low (51.3%)	WJ letter identification	77.43			
	WJ word attack	84.08			
	TOWRE SWE	78.28			
	TOWRE PDE	78.60			
	GORT fluency	4.96			
	DIBELS ORF	7.81			
	CTOPP phonological	87.80			
	CELF USP	6.72			
	QRI listening comp	2.39			
Low-average (25.7%)	WJ letter identification	92.71	15.28		
	WJ word attack	101.50	17.42		
	TOWRE SWE	93.57	15.29		
	TOWRE PDE	91.67	13.07		
	GORT fluency	8.08	3.12		
	DIBELS ORF	27.62	19.81		
	CTOPP phonological	99.00	11.20		
	CELF USP	8.50	1.78		
	QRI listening comp	3.06	0.67		
Average (17.9%)	WJ letter identification	107.79	30.36	15.08	
	WJ word attack	112.61	28.53	11.11	
	TOWRE SWE	111.61	33.33	18.04	
	TOWRE PDE	106.42	27.82	14.75	
	GORT fluency	10.39	5.43	2.31	
	DIBELS ORF	55.82	48.01	28.20	
	CTOPP phonological	111.69	23.89	12.69	
	CELF USP	10.69	3.97	2.19	
	QRI listening comp	3.99	1.60	0.93	
Above average (5.1%)	WJ letter identification	117.81	40.38	25.10	10.02
	WJ word attack	118.76	34.68	17.26	6.15
	TOWRE SWE	120.59	42.31	27.02	8.98
	TOWRE PDE	113.46	34.86	21.79	7.04[a]
	GORT fluency	12.10	7.14	4.02	1.71
	DIBELS ORF	81.42	73.61	53.80	25.60
	CTOPP phonological	111.55	23.75	12.55	−0.14[a]
	CELF USP	12.34	5.62	3.84	1.65
	QRI listening comp	3.93	1.54	0.87	−0.06[a]

Note. WJ = Woodcock Johnson; TOWRE = Test for Word Reading Efficiency; SWE = sight word efficiency; PDE = phonemic decoding efficiency; GORT = Gray Oral Reading Test; DIBELS ORF = Dynamic Indicators of Basic Early Literacy Skills–Oral Reading Fluency; CTOPP phonological = phonological processing composite; CELF USP = Clinical Evaluation of Language Fundamentals–Understanding Spoken Paragraphs subtest; QRI = Qualitative Reading Inventory–5.
All differences significant at $p < .05$ except [a] *ns*.

evident. This highlights the importance of investigating data sets beyond sample-level means as a more nuanced understanding of the heterogeneity of the sample has important implications for targeted reading instruction. Once the distinct reading trajectories were established, they were further analyzed to determine whether subcomponent skills of reading comprehension differentiated the groups. Significant differences were noted across the majority of the subcomponent skills—both code- and meaning-related skills—across groups. These findings are discussed in the following in more detail, as well as their implications for instructional practice.

Growth in Reading Comprehension by Latent Class

Four distinct groups of students were identified based on their reading comprehension trajectories during the first-grade

year. Growth across the first-grade year occurred for all groups except the *Above Average* group; however, the rank order of the groups remained the same from fall to spring. That is, no group surpassed any other group between fall and spring in terms of reading comprehension. Yet there was a notable finding regarding the *Average* and *Above Average* groups. The *Average* group scored similarly to the *Above Average* group by spring although it scored substantially lower in fall. The raw GORT reading comprehension scores for each of these two groups corresponded to scaled scores in the average range. Together, these groups may represent different levels of typical development. It should be noted that, even in the fall of the first-grade year, the *Above Average* group, representing just 5% of the sample, was performing approximately 1 *SD* above average on the reading comprehension measure, but this did not continue into spring; their spring score corresponded to average performance. The *Average* group, on the contrary, scored approximately average in fall and was also average in spring. This might suggest that the *Above Average* group regressed to the mean by spring. Alternatively, it may be that teachers did not provide any specialized or targeted instruction to the *Above Average* group as they did not appear to be experiencing any risk based on fall performance.

By spring testing, three of the four groups reached at least the average range on the reading comprehension assessment; combined, this represented 49% of the sample scoring in the average range compared with 75% at the beginning of the school year. The *Low* group, which accounted for 51% of the sample, did not reach the average range on reading comprehension and remained between 1 and 2 *SD*s below average at the end of first grade. It is cause for concern that just more than one half of the sample scored below average at the end of first grade. Previous literature indicates that the first-grade year is extremely important for long-term attainment in reading (Cunningham & Stanovich, 1997). Thus, the *Low* group represents a sizable proportion of students who may be at long-term risk if they do not receive targeted instruction that can dramatically increase their reading comprehension achievement. Students in this group should be relatively easy for educational practitioners to identify as their scores across all subcomponent skills were more than a standard deviation below average (except phonological skills, which were near the cutoff between average and below average). This skill profile suggests instruction that is solely focused on word reading, and decoding is likely not sufficient to ameliorate later reading comprehension difficulties if this group continues to struggle with meaning-related skills. Rather, this group appears to require a comprehensive curriculum that complements code-related skills with meaning-related instruction.

Differences on Subcomponent Skills Across Groups

Great variation was seen across groups on the reading subcomponent skills; the *Low, Low-Average*, and *Average* classes were significantly different on all subcomponent skills. In addition, the *Low* and *Low-Average* classes were also significantly different on all subcomponents compared with the *Above Average* class. This finding supports emerging literature that have investigated the heterogeneous nature of reading development (e.g., Boscardin et al., 2008; Grimm et al., 2018) and demonstrates the wide range of reading development during the first-grade year. Of particular importance is the finding that skill development differed across both subcomponent skills related to decoding skills and those that are more meaning-based such as listening comprehension. From an instructional standpoint, these findings highlight the wide range of reading instruction necessary to adequately address the reading needs of students in first grade.

When comparing the *Average* and *Above Average* groups, significant differences were not noted on phonological skills. From a developmental perspective, this makes sense, considering that these students were performing average on reading comprehension assessments; their phonological skills were well-developed to enable more accurate decoding and reading fluency. Whereas phonological skills were comparable between these two groups, significant differences were noted on three of the four word reading measures although the groups were comparable by the end of first grade on reading comprehension. That is, the differences in word reading measures were likely related to differences in reading comprehension in the fall of first grade, but not by spring as these two groups' reading comprehension scores converged by then. This finding may indicate that, as students develop more enhanced higher order reading skills, such as reading comprehension, the relation between decoding and comprehension may be less impactful. Importantly, the finding converges with prior empirical evidence suggesting that word reading skills are constrained in that their effects are greatest during early stages of the development of reading comprehension and become less pronounced as children progress (e.g., Hoover & Gough, 1990; Kershaw & Schatschneider, 2012). Significant differences between the *Average* and *Above Average* groups were not seen on the QRI listening comprehension measure. There are two important points to make about this finding. First, The QRI is not a norm-referenced assessment, so this finding may be a reflection of the measure. However, from a skill development perspective, it would make sense that students who perform equally as well on comprehension measures by the end of first grade demonstrate similar

strengths in listening comprehension. Yet the fact that the *Average* and *Above Average* groups scored similarly on CELF USP, a different—and standardized—measure of linguistic comprehension, cannot be ignored. Table 4 illustrates that there was consistency in terms of the differences on QRI and CELF USP with respect to the rank-ordering of the latent classes. That is, latent classes that demonstrated lower reading comprehension achievement also demonstrated lower linguistic comprehension achievement on both measures although the difference in CELF USP was nonsignificant between the *Above Average* and *Average* groups. However, the CELF USP scores are scaled, and a close examination reveals that the difference between the *Above Average* and *Average* groups is only approximately one half of a standard deviation. Both groups scored within the average range although the scores were statistically significantly different. Thus, it is unlikely that this difference would raise concerns among educational practitioners.

The results of this study demonstrate a more nuanced understanding of early reading comprehension development and indicate that reading comprehension is not solely driven by word reading and decoding differences, but instead by several different subcomponent skills, with significant differences noted between most groups across all measures. These findings have important implications for instructional practice

Implications for Practice

From a broad perspective, the findings in this study highlight the importance of understanding the nuanced development of reading during the first-grade year. A more in-depth understanding of reading skill profiles should be utilized to inform more effective and targeted reading instruction, especially for children whose reading acquisition is impaired in the early grades. Although instruction in more basic word reading and decoding is essential for students who are struggling to develop early reading skills, some consideration of how to also enhance meaning-related comprehension skills is necessary. General education teachers, reading specialists, and special education teachers should be made aware of these differences seen across children and, when relevant, should allocate some time for developing meaning-related skills in students performing below average in reading comprehension. It is possible that this is particularly important for students who come from lower SES backgrounds as more than 60% of the sample in this study qualified for free and reduced-price lunch.

Sometimes, the most difficult group to make instructional decisions about are not the lowest performing students, who in most cases are very clearly in need of supplemental reading instruction, but, instead, the students who are performing closer to average but still may warrant some additional targeted instruction; in this study, this is the *Low-Average* class. These students, representing just more than 25% of the sample, most likely require some additional targeted instruction tailored to their subcomponent skill reading deficits to ameliorate their early reading risk. In addition, their progress in these skill areas should be regularly monitored to ensure that they are not falling further behind the students who are performing at or above average.

Limitations and Future Directions

Approximately, half of the sample was classified into the lowest performing group, which met criteria for at-risk status, and this may be a larger proportion of students than expected. However, this may be due to a sampling issue. The majority of the sample comprised students from racial and ethnic minority and low-SES backgrounds, which are known to be risk factors for potential reading difficulties (McCoach et al., 2006; Morgan et al., 2008). While the proportion of at-risk students may not align with other research samples, the findings highlight the need for reading resources to be made available to these populations of students.

This study provides multiple avenues for future studies to pursue. First, the findings need to be replicated with independent samples. Future studies should also investigate the stability of these subgroups beyond first grade. The *Average* group nearly caught up with the *Above Average* group by the end of first grade, but it is not clear whether the reading performances of the *Average* and *Above Average* groups further align in later grades. Related, although it was clear that the *Low-Average* and *Low* groups did not catch up to the higher performing groups, further investigation is needed to identify which subcomponent skills might be most amenable to intervention. Although this study demonstrated that the latent growth classes could be differentiated by subcomponent skills of reading comprehension, it did not identify a hierarchy of subcomponent skills in terms of their contributions to reading comprehension. However, a unique contribution of this study was to identify the importance of meaning-related skills. While word reading intervention in early elementary years has been identified as beneficial, it is not yet clear whether intervention in meaning-related skills in early elementary years can ameliorate future reading difficulties. The findings from this study suggest that this may be an important avenue for future reading studies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the Institute of Education Sciences, U.S. Department of Education, grant R324A150091 to the University of Virginia. Alyssa R. Henry received support during the drafting of this manuscript from a Postdoctoral Research Fellow Training Program Grant from the National Center for Special Education Research at the Institute of Education Sciences (R324B180034). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences or the U.S. Department of Education.

Supplemental Material

Supplemental material for this article is available on the *Journal of Learning Disabilities* website with the online version of this article.

References

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*(4), 332–345. https://doi.org/10.1177/0022219410369067

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, *19*(9), 933–958. https://doi.org/10.1007/s11145-006-9024-z

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes*, *21*(2), 1–22.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, *43*(1), 272–311. https://doi.org/10.1177/0081175012470644

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, *23*, 20–31. https://doi.org/10.1080/10705511.2014.955104

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*, 3–27. https://doi.org/10.1093/pan/mph001

Boscardin, C. K., Muthe´n, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, *100*, 192–208. https://doi.org/10.1037/0022-0663.100.1.192

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

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

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, *3*(4), 331–361. https://doi.org/10.1207/s1532799xssr0304_2

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32*(1), 38–50. https://doi.org/10.1044/0161-1461(2001/004)

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

Chard, D. J., Vaughn, S., & Tyler, B.-J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, *35*(5), 386–406. https://doi.org/10.1177/00222194020350050101

Clark, S., & Muthén, B. (2009). Relating Latent Class Analysis results to variables not included in the analysis. http://www.stat model.com/download/relatinglca.pdf

Compton, D. L., Fuchs, D., Fuchs, L. S., Elleman, A. M., & Gilbert, J. K. (2008). Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. *Learning and Individual Differences*, *18*, 329–337. https://doi.org/10.1016/j.lindif.2008.04.003

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945.

Eason, S. H., Sabatini, J., Goldberg, L., Bruce, K., & Cutting, L. E. (2013). Examining the relationship between word reading efficiency and oral reading rate in predicting comprehension among different types of readers. *Scientific Studies of Reading*, *17*(3), 199–223. https://doi.org/10.1080/10888438.2011.652722

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, *3*(1), 30–37. https://doi.org/10.1111/j.1750-8606.2008.00072.x

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

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, *88*(1), 3–17. https://doi.org/10.1037/0022-0663.88.1.3

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

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

Grimm, R. P., Solari, E. J., McIntyre, N. S., & Denton, C. A. (2018). Early reading skill profiles in typically developing

and at-risk first grade readers to inform targeted early reading instruction. *Journal of School Psychology*, *69*, 111–126. https://doi.org/10.1016/j.jsp.2018.05.009

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology*, *16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441

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

Juel, C., & Leavell, J. A. (1988). Retention and nonretention of at-risk readers in first grade and their subsequent reading achievement. *Journal of Learning Disabilities*, *21*(9), 571–580. https://doi.org/10.1177/002221948802100910

Kendeou, P., Savage, R., & van den Broek, P. (2009a). Revisiting the simple view of reading. *British Journal of Educational Psychology*, *79*(2), 353–370. https://doi.org/10.1348/978185408X369020

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

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

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology*, *102*(3), 652–667. https://doi.org/10.1037/a0019643

Leslie, L. & Caldwell, J. (2011). *Qualitative reading inventory*. Pearson Education.

Masyn, K. E. (2013). Latent class analysis and finite mixture modelling. In T. D. Little (Ed.), *Oxford Library of Psychology. The Oxford handbook of quantitative methods: Statistical analysis* (551–611). Oxford University Press.

Mather, N., & Jaffe, L. E. (2016). *Woodcock-Johnson IV: Reports, recommendations, and strategies*. John Wiley.

Mathes, P. G., & Denton, C. A. (2002). The prevention and identification of reading disability. *Seminars in Pediatric Neurology*, *9*(3), 185–191. https://doi.org/10.1053/spen.2002.35498

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*(2), 148–182. https://doi.org/10.1598/RRQ.40.2.2

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*(1), 14–28. https://doi.org/10.1037/0022-0663.98.1.14

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, *71*(4), 445–463. https://doi.org/10.1177/001440290507100404

Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly*, *31*(4), 187–198. https://doi.org/10.2307/25474651

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the SEM algorithm. *Biometrics*, *55*(2), 463–469. https://doi.org/10.1111/j.0006-341X.1999.00463.x

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

National Institute of Child Health and Human Development (NICHD). (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: Reports of the subgroups* (NIH Publication No. 00-4754). U.S. Government Printing Office.

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*(4), 443–468. https://doi.org/10.1080/01690960344000008

Raykov, T., & Marcoulides, G. A. (2006). On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling*, *13*(1), 130–141. https://doi.org/10.1207/s15328007sem1301_7

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research*, *95*(5), 259–272. https://doi.org/10.1080/00220670209596600

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development*, *61*(6), 1728–1743. https://doi.org/10.1111/j.1467-8624.1990.tb03562.x

Schwartz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Semel-Mintz, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). Psychological Corporation.

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading*, *17*(2), 108–133. https://doi.org/10.1080/10888438.2011.618153

Skibbe, L. E., Grimm, K. J., Stanton-Chapman, T. L., Justice, L. M., Pence, K. L., & Bowles, R. P. (2008). Reading trajectories of children with language difficulties from preschool through fifth grade. *Language, Speech, and Hearing Services in Schools*, *39*(4), 475–486.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S., Smith, L., & Dianda, M. (1996). Success for All: A summary of research. *Journal of Education for Students Placed at Risk*, *1*(1), 41–76. https://doi.org/10.1207/s15327671espr0101_6

Solari, E. J., Grimm, R. P., McIntyre, N. S., & Denton, C. A. (2018). Reading comprehension development in at-risk vs. not at-risk first grade readers: The differential roles of listening comprehension, decoding, and fluency. *Learning and Individual Differences*, *65*, 195–206. https://doi.org/10.1016/j.lindif.2018.06.005

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947. https://doi.org/10.1037//0012-1649.38.6.934

Tilstra, J., McMaster, K., Van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, *34*(1), 33–58. https://doi.org/10.1177/002221940103400104

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161–188). Routledge.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency-Second Edition (TOWRE-2)*. Pro-Ed.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. https://doi.org/10.1177/0022219411432685

Vellutino, F. R., Scanlon, D. M., & Reid Lyon, G. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers. *Journal of Learning Disabilities*, *33*(3), 223–238. https://doi.org/10.1177/002221940003300302

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., . . . Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468–479.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *CTOPP-2: Comprehensive test of phonological processing*. Pro-ed.

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray Oral Reading Test–Fifth Edition (GORT-5): Examiner's manual*. Pro-Ed.