Read Writ (2007) 20:885–907
DOI 10.1007/s11145-007-9058-x

Clusters of second and third grade dysfluent urban readers

Margaret E. Pierce · Tami Katzir · Maryanne Wolf · Gil G. Noam

Published online: 20 June 2007

Abstract This study examined the heterogeneity within a sample of 140 urban second and third graders identified as at-risk for reading failure due to inefficient word and/or nonword reading. Cluster analyses were conducted using standardized factor scores from a four-factor structural equation model characterizing reading performance in this sample. These standardized factor scores represented performance on four distinct factors: efficiency of word-level skills, text level skills, decoding, and vocabulary. Results identified four clusters of children who show distinctive patterns of performance on the four factors of reading. These under-studied groups show different compositions along demographic categories and reading disability categories. Additionally, they have unique instructional needs that call for differentiated instruction in the domains of phonological decoding, fluency, text reading, and vocabulary.

Keywords Cluster analysis · Reading disability · Reading fluency · Subtypes · Urban children · Vocabulary

M. E. Pierce (✉)
Student Development and Pupil Personnel Services, University of Massachusetts Amherst, 166 Hills House South, 111 Infirmary Way, Amherst, MA 01003, USA
e-mail: mpierce@educ.umass.edu

T. Katzir
Department of Learning Disabilities, University of Haifa, Haifa 31905, Israel

M. Wolf
Eliot-Pearson Department of Child Development, Tufts University, 105 College Avenue, Medford, MA 02155, USA

G. G. Noam
Program in Education, Afterschool & Resiliency, Harvard Medical School and McLean Hospital, Mill Street Lodge, 115 Mill Street, Belmont, MA 02478, USA

Introduction

Since the *No Child Left Behind Act of* 2001 (2001) was passed, public attention has focused on the persistent failure of America's urban schools to improve the reading proficiency of disadvantaged groups, including poor children, minority children, and English language learners (ELLs). Many urban children show a range of reading difficulties beyond those typically associated with reading disability due to the compounding effects of English language proficiency (August & Hakuta, 1997), lack of familiarity with Standard English (Charity, Scarborough, & Griffin, 2004; Labov, 1995), and characteristics of the environment associated with low socioeconomic status (SES) (Duncan & Seymour, 2000; for a review see Snow, Burns, & Griffin, 1998). Due to the myriad of factors affecting the trajectories of reading development among urban children, there is considerable heterogeneity among failing readers in urban schools, though it is unclear how early this heterogeneity emerges. To help remediate the reading difficulties exhibited by these struggling students, teachers in urban schools are faced with the task of deciding which instructional methods to use with their students based on a careful consideration of each child's profile of strengths and weaknesses. Thus the delineation of valid profiles is critical for the selection of appropriate instructional methods and materials.

There is little existing research that can assist in the delineation of these profiles for struggling readers attending urban schools. One study, summarized below, offers a beginning look at the variation in reading development among upper elementary struggling readers attending urban schools. To build upon this research, this study was designed to characterize literacy profiles of younger urban children struggling to acquire beginning reading skills. Converging results from studies such as these will enrich our understanding of the heterogeneity among urban struggling readers across development, with implications for methods of selecting and implementing more individualized quality reading instruction suited to different profiles of strength and weakness.

Subtypes of fourth grade readers below the bar

Riddle Buly and Valencia (2002; Valencia & Riddle Buly, 2004) examined the heterogeneity among a sample of students in a Northwestern urban school district who had failed the fourth grade state reading test. Their results offer a first look at subtypes of children showing unique patterns of performance on factors of reading, and their methodology provides a tool for the expansion of this examination to other populations of urban children. Using a battery of literacy measures, Riddle Buly and Valencia (2002) identified three correlated factors underlying the reading performance of their failing fourth grade students. These factors represented variation in word identification skills, fluency (including rate and expression), and meaning (including comprehension and vocabulary). The authors identified at least six subgroups of children showing unique profiles along these three factors of reading: automatic word callers, struggling word callers, word stumblers, slow comprehenders, slow word callers, and disabled readers. In a paper discussing these results, Valencia and Riddle Buly (2004) reiterated the need to conduct diagnostic assessments (in addition to state tests) to identify the specific nature of each failing student's needs, the need for multilevel, flexible, small-group instruction, and the need for continued language support for children who had exited English as a Second Language programs but whose limited English proficiency still interfered with their reading comprehension.

The reading profiles described by Valencia and Riddle Buly (2004) offer a compelling argument for flexible and individualized instruction in urban schools. The heterogeneity they described is consistent with research on the reading development of upper elementary school children, when fluency and upper level comprehension skills become increasingly important in predicting outcomes (Snow et al., 1998). However, it remains unclear whether similar patterns of heterogeneity characterize the reading development of younger, urban children at risk for reading failure. Perhaps in earlier stages of reading development, when the primary task is learning to identify words, struggling students show a more homogeneous profile. An investigation of this kind, which is the purpose of this study, must begin with a model of the factor structure underlying reading performance of younger children.

The factor structure of reading performance

In a previous study (Pierce, Katzir, Wolf, & Noam, 2006), we investigated the factor structure of reading performance among second and third grade urban readers at risk for reading failure, using structural equation modeling. This model offers a starting place for an examination of heterogeneity in this sample inspired by Valencia and Riddle Buly's methodology. The findings from this study offered an expansion of unidimensional models of reading among urban children (Mehta, Foorman, Branum-Martin, & Taylor, 2005), by characterizing the second order factor structure of reading performance in urban at-risk readers. Based on data from a battery of common literacy measures administered to 140 second and third graders identified as dysfluent at the word level, these results showed that variation in reading performance is best captured by a four-factor structural equation model that parallels Adams' (1990) four-component model of skilled reading. In the sample of struggling second and third grade readers, reading performance varied along four correlated dimensions: word-level efficiency, text level skills, decoding, and vocabulary. Although correlated, the emergence of four unique factors suggests that vocabulary, decoding, efficiency, and text level skills are developmentally distinct among struggling second and third grade urban children. A more nuanced characterization of these factors is permitted by examining the relative strength of the measured variables that serve as indicators of each factor.

The efficiency factor was predicted by five measured variables, with sight word efficiency emerging as the strongest indicator of this factor, followed by rapid letter naming, passage reading rate, phonemic decoding efficiency, and knowledge of words with multiple meanings. Because sight word efficiency emerged as the marker variable, we interpreted this factor as emphasizing efficiency with word-level skills. Measured variables assessing receptive vocabulary knowledge and expressive knowledge of multiple meanings served as the strongest indicators of the vocabulary factor. Passage reading comprehension also served as a weak but significant indicator of the vocabulary factor, consistent with research highlighting the importance of vocabulary knowledge in supporting reading comprehension (Anderson & Freebody, 1981; Kameenui, Carnine, & Freshi, 1982). The decoding factor was predicted by four of the measured variables, with untimed nonword reading emerging as the marker variable for this factor. Additional indicators of the decoding factor included phoneme elision, phonemic decoding efficiency, and passage reading accuracy. Finally, the text level skills factor was predicted by three of the measured variables, with passage reading accuracy emerging as the strongest indicator of this factor, followed by passage reading comprehension and passage reading rate.

Comparing the factor structure models across age

A comparison of the factor structure model underlying the performance of fourth graders who failed a high stakes reading test and the model underlying the performance of second and third grade dysfluent readers highlights several interesting similarities and differences. Both structural equation models identified an efficiency factor, suggesting that efficiency is an important source of variation among early *and* late elementary school students who are struggling with reading. However, while Riddle Buly and Valencia (2002) reported that reading comprehension and vocabulary loaded with similar magnitude on their meaning factor, the results from our study suggest that comprehension is more weakly related to vocabulary among second and third grade struggling readers. This may be related to differences in the proportion of low frequency vocabulary words included in texts at the second, third, and fourth grade level. Another difference is apparent in the factor structure of reading accuracy. Among fourth grade failing readers, passage accuracy and nonword reading loaded onto a unitary factor, which Riddle Buly and Valencia called ''Word Identification.'' However, with younger struggling readers, variation in accuracy consists of two distinct factors, text level skills (indicated best by passage reading accuracy), and decoding (indicated best by untimed nonword reading). This finding suggests that the translation of word-level skills to text reading is not automatic among beginning readers.

Aims of the present study

The efforts of those behind the *No Child Left Behind* legislation have focused a national commitment to providing quality reading instruction to all children. However, our ultimate success in this endeavor relies on the extent to which we can provide tailored instruction to subtypes of children showing different literacy profiles. While previous research has contributed to our understanding of what these subtypes look like in an upper elementary school urban population, further research is needed to characterize literacy profiles of younger urban children struggling to acquire beginning reading skills. The purpose of this study was to identify subgroups of second and third grade children who show unique patterns of performance and who have unique instructional needs. The goal, therefore, was to identify a cluster solution with sufficient validity, reliability, and coverage to characterize the heterogeneity of skills within a sample of 140 second and third grade urban children identified as at risk for reading failure. Specifically, the study was designed to answer the following research questions:

1. Does the pattern of performance suggested by group averages on four factors of literacy fit the majority of participants, or are there different, compelling patterns of performance in the sample?
2. What are the literacy profiles exhibited by different clusters of children, and what implications do these profiles have for intervention?
3. How do the literacy profiles exhibited by different clusters of children relate to current models of reading disability?

Method

Participants

Participants were 140 second and third graders enrolled in five well-established school-based afterschool programs in two metropolitan areas, Phoenix and Boston. Subjects were recruited to participate in a study of a reading intervention and a socio-emotional intervention taking place in the afterschool programs (Noam, Wolf, & Katzir, 2003). Selection criteria were designed to identify children at risk for reading failure. Accordingly, participation was offered only to children who scored more than two-thirds of a standard deviation below the mean on one of the subtests or the composite from the Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999). Children who had been retained in first or second grade, and children who exhibited severe hearing, visual, or neurological impairments were excluded from the study.

While descriptive data on this sample is available elsewhere (Pierce et al., 2006), it is important to note that this sample of children was selected to display inefficiencies in word-level skills. This selection criterion is wider than the more classic method of identifying children whose untimed word reading is poor, indicating deficits in word reading accuracy. In addition to inaccurate readers, this sample includes children whose word or nonword reading is accurate but slow. However, this sample may not include the full range of children who display difficulties comprehending grade level text, particularly if those difficulties stem from deficits in listening comprehension rather than decoding ability (Gough & Tunmer, 1986).

Of the 140 participants who entered the study, 69 children were in second grade and 71 children were in third grade. Fifty-eight participants attended afterschool in the Boston area, while 82 participants attended school in the Phoenix area. There were 86 male and 54 female participants in the study. This sample was linguistically, ethnically, and economically diverse. According to questionnaire data provided by parents, this sample was 60% Hispanic/Latino, 49% white, 12% black or African American, and 4% Asian. Additionally, 49% of the children spoke

English (and only English) as their first language, while 46% of the children spoke a language other than English (and only that language) as their first language. Finally, 89% of the mothers or female guardians and 84% of fathers or male guardians earned less than $35,000 a year.

Instructional context

All five schools serving the children in this sample use a systematic phonics program as part of the second and third grade language arts curriculum (including ZooPhonics, Project Read: Phonology Strand, and Wilson Fundations). However, there was considerable variation in the nature of the reading programs used, with some classrooms using whole class basal instruction (McGraw Hill: Treasures) and other classrooms using Guided Reading with flexible groupings of children (Learn to Read by Reading, Rigby PM Readers). Additionally, it is important to note that after the collection of these data, many of the children participated in a multi-componential reading fluency intervention (RAVE-O) and a socio-emotional intervention (RALLY) administered in afterschool programs.

General procedure

For these analyses age-based standard scores from literacy measures administered during the screening and pretesting phases of data collection were used. Screening and pretesting took place during the school day at the five participating elementary schools. All children were screened during the first 2 months of the school year, and all children were pretested during the first 4 months of the school year. In all cases, pretesting was completed before the child's participation in the afterschool intervention commenced.

A battery of literacy measures was administered to all 140 subjects to assess performance on a number of literacy tasks. The test battery included measures shown to be indicators of the four factors of reading: decoding, word-level efficiency, text level skills, and vocabulary.

Phonological processing

Phonological awareness was assessed using the Elision Subtest of the Comprehensive Test of Phonological Processes (CTOPP: Wagner, Torgesen, & Rashotte, 1999), which requires the child to isolate syllables or phonemes within spoken words. Early items require the child to remove words from compound words (e.g., ''say 'popcorn' without saying 'pop''), while later items require the child to remove individual phonemes from within a word (e.g., ''say 'strain' without saying/r/'').

Vocabulary knowledge

Expressive vocabulary knowledge was assessed using the Multiple Meanings subtest from The Word-R Test (Huisingh, Barrett, Zachman, Blagden, & Orman, 1990), which measures the ability to express more than one meaning for each of 15 words (e.g., WATCH: you wear it on your wrist and it tells you what time it is OR you look at someone with your eyes). All children were asked to define all 15 words. The raw score reflects the number of words for which the child could provide at least two correct definitions.

Receptive vocabulary knowledge was assessed using The Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III: Dunn & Dunn, 1997), which measures the ability to recognize the proper meaning of a word. On the PPVT, the child is asked to select the picture (from four choices) that best represents the meaning of an orally presented word.

Speed of lexical retrieval

Speed of lexical retrieval for letter names was assessed using the Letter Naming subtest from the Rapid Automatized Naming Test (RAN: Wolf & Denckla, 2004). This test measures the speed and accuracy of a child's retrieval of letter names. Subjects were asked to name stimulus letters arranged in an array (five rows, 10 items per row) as quickly and accurately as possible. This version of the RAN includes the following letters in the array: a, o, s, p, and d.

Nonword decoding

The Word Attack subtest from the Woodcock Reading Mastery Test (WRMT-R: Woodcock, 1997) assesses skill in using phonic and structural analysis to read nonsense words of increasing complexity, beginning with ''dee'' and continuing through ''pnomocher.'' While not officially timed, subtests from the WRMT require the child to provide a natural reading of the word within about 5 s.

Word-level fluency

Speed (and accuracy) for reading words and nonwords was assessed with the Test of Word Reading Efficiency (TOWRE: Torgesen et al., 1999). The Sight Word Efficiency subtest contains 104 real words of increasing difficulty arranged in four columns. The subject is required to read aloud as many words as possible within the space of 45 s. The Phonemic Decoding Efficiency subtest contains 63 increasingly difficult nonwords arranged in three columns. The subject is required to read as many nonwords as possible within the space of 45 s.

Oral reading fluency (connected-text)

The Gray Oral Reading Test, 4th Ed. (GORT-4: Wiederholt & Bryant, 2001) measures the accuracy and rate with which children read connected texts. The GORT consists of 14 passages of increasing difficulty and length. As the child reads each passage aloud, the examiner takes careful note of the time required to read the passage and the number of deviations from print. An accuracy score reflects the number of deviations from the text during oral reading, including substitutions, omissions, insertions, self-corrections, and repetitions. A rate score reflects the number of seconds required to read each passage.

Oral reading comprehension

Oral reading comprehension was assessed with the comprehension index from the Gray Oral Reading Test, 4th Ed. (Wiederholt & Bryant, 2001). After reading the passages (described previously), the child is read five multiple-choice questions about the content of each passage. Comprehension questions on the GORT require direct recall (e.g., ''What was the man holding?''), inference (e.g., ''How do you think the girl felt when she saw the surprise?''), and synthesis (e.g., ''What is the best name for this story?''). It should be noted that the reading comprehension subtest of the GORT may overestimate reading comprehension skill as many of the items can be answered correctly without reading the passage (Keenan & Betjemann, 2006).

Analyses

Preliminary statistical checks were conducted to ensure that univariate and multivariate outliers were not present. To investigate the first research question, the K-Means Cluster Analysis procedure was used to generate cluster solutions for evaluation. This procedure begins by using the factor values from the first *k* cases as temporary estimates of the *k* cluster means, where the number of clusters (*k*) is determined by the researcher. Through an iterative process, cases are added to the clusters and the cluster centers are recalculated (Norusis, 2006). With cluster analytic techniques, each cluster in the final solution must display a unique profile that appears to be theoretically relevant within the population being examined. Additionally, each cluster must consist of a sufficient number of cases to ensure that outlying cases are not exerting undue influence on the cluster solution.

Cluster analysis was conducted using standardized factor scores generated from factor score weights produced from a four-factor structural equation model characterizing reading performance in this sample (for a description of this method, see Pierce et al., 2006). These standardized factor scores represented performance on four distinct factors: efficiency of word-level skills, text level skills, decoding, and vocabulary. Two-, three-, and four-cluster solutions were estimated. (Five-, six-, and seven-factor solutions were also requested, but the solutions were found to be unstable using the split sample method described below.) The two-cluster solution generated two groups, undifferentiated on vocabulary (with vocabulary factor scores at the sample mean). The first group showed comparatively low scores on decoding, efficiency, and text level skills, while the second group showed comparatively high scores on these three factors. The three-cluster solution generated three distinct groups, two with comparatively low scores on vocabulary and one with a comparatively high score on vocabulary. The high vocabulary group also showed a relative strength in decoding and a relative weakness in text level skills. The first low vocabulary group showed a relative weakness in efficiency, text level skills, and decoding, while the second low vocabulary group showed a relative strength in the remaining three factors. The four-cluster solution generated two low vocabulary clusters and two high-vocabulary clusters, differentiated along decoding, efficiency, and text reading. Given the demonstrated importance of decoding, fluency, and vocabulary (NRP, 2000) in reading proficiency and instruction, we identified this solution as having the greatest theoretical relevance. The results of this cluster solution are presented below.

To address the second research question, descriptive statistics for the marker variables and select demographic variables were examined for each of the clusters. Because cluster analysis maximized differences among the groups on the four factors, analyses of variance on the marker variables (which are highly correlated with the factors) would be inappropriate. Rather, visual inspection of each cluster's profile was compared to the total group profile along two dimensions: shape of the profile and altitude of the profile. Additionally, chi square analyses on the demographic data were conducted except where analyses violated the assumptions regarding minimum expected cell frequency (e.g., too few expected families from nonlow income homes in each cluster).

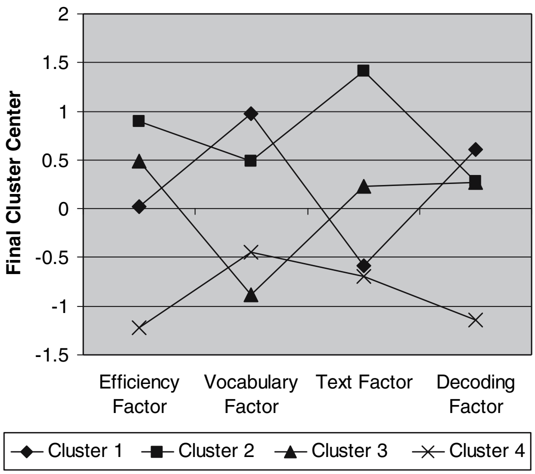
Finally, to address the third research question, chi square analyses were conducted to determine whether cluster membership was related to categorization as no deficit, phonological deficit, naming speed deficit, or double deficit, except where analyses violated the assumptions regarding minimum expected cell frequency (e.g., too few expected double deficit and single phonological deficit children in each cluster). For these analyses, participants were categorized as showing a *naming speed deficit* if they earned a RAN Letters standard score more than one standard deviation below the norm sample mean (i.e., <85) and an Elision standard score within or above the average range. Participants were categorized as showing a *phonological deficit* if they earned an Elision standard score more than one standard deviation below the norm sample mean (i.e., <7) and a RAN Letters score within or above the average range. Participants were categorized as showing a *double deficit* if they earned standard scores on both RAN Letters and Elision that were more than one standard deviation below the norm sample mean. Finally, participants were categorized as showing *no deficit* if they earned scores on both RAN Letters and Elision within or above the average range.

Results

Research question 1: groups of children showing unique profiles along four factors

The results of the four-cluster solution are presented in Fig. 1, which shows the cluster centers for the four factors for each of the four clusters. Because the factor scores were standardized using z scores, the cluster centers offer an intra-sample comparison but not a comparison to the standardized tests' norming samples. As this figure shows, Cluster 1 (N = 38) showed comparatively high scores on the vocabulary and decoding factors and comparatively low scores on the word-level

Fig. 1 Four-cluster solution distinguishing subgroups of at-risk urban readers



efficiency and text level skill factors. Cluster 2 ($N = 27$) showed relatively high scores on the word-level efficiency, text level, and vocabulary factors, and a decoding score near the sample average. Cluster 3 ($N = 39$) showed comparatively low scores on the vocabulary factor, with scores on the remaining three factors at or above the sample means, while cluster 4 ($N = 36$) earned low scores on all four factors. To evaluate the internal consistency of cluster membership within this four-cluster solution, the sample was randomly divided in half. Each half ($N = 70$) was then clustered using the K-means procedure. The final cluster centers from half A were used to classify the cases from half B, and the final cluster centers from half B were used to classify the cases from half A. This process resulted in two sets of cluster membership assignments, which were compared by calculating Cohen's kappa ($\kappa = 0.562$, SE = .052), a statistic that captures the agreement between two raters (in this case, two independent attempts at categorization into four categories). This kappa value is within the range characterized as *moderate* by Landis and Koch (1977), and within the range characterize as *fair* by Fleiss (1981). Thus the cluster solution was judged to be internally consistent.

The first research question addressed whether the pattern of performance suggested by group averages on four factors of literacy fit the majority of subjects, or whether different, compelling patterns of performance in the sample emerged. The factor values for each cluster are represented graphically in Fig. 1. Because the factor scores are presented as z scores, the total group average for each factor is zero. In Fig. 1, the profile exhibited by the total group is represented by the line extending horizontally through zero. Because the clusters were generated to be maximally distinct, F tests are an inappropriate test of the hypothesis that that cluster means are equal. Rather, each cluster's profile was compared to the total group profile along two dimensions: shape of the profile and altitude of the profile.

An examination of the profile and altitude of the trend lines in Fig. 1 shows that none of the four clusters shows a pattern of performance similar to the total group averages. The shape of the profile exhibited by Cluster 4 (see Fig. 1) is relatively flat, suggesting that this cluster is the most similar to the profile suggested by the sample means (represented by the line extending horizontally through zero in Fig. 1). However, Cluster 4's mean factor scores are all well below the total sample's means, suggesting that this group earned lower scores on the measures, particularly those contributing to the efficiency, text reading, and decoding factors. An examination of the differences in profile shape in Fig. 1 also shows that Cluster 1 shows a relative strength in vocabulary and a relative weakness in text level skills as compared to the total group. Cluster 2 shows a relative strength in text level skills, and Cluster 3 shows a relative weakness in vocabulary. These results suggest that using the sample averages to characterize performance and design intervention would fail to acknowledge the specific needs of each cluster.

Research question 2: characterization of the reading profiles exhibited by four clusters

The means and standard deviations of marker variables representing each factor for each of the four clusters and the total sample are presented in Table 1. Because the marker variables are highly correlated to the factors, which were the basis of the assignment of cases to the clusters, analyses of variance on cluster marker variable scores would be redundant. Nevertheless, the marker variable mean scores allow a comparison to the age-matched norming group from these norm-referenced tests. A description of each cluster is presented below.

Cluster 1

Cluster 1 contained the highest proportion of second graders (63%), of English monolinguals (68%), of children from higher SES families (32%), and of children categorized as displaying a single naming speed deficit (32%). As Table 2 shows, Cluster 1 showed a mean standard score at the 50th percentile on Word Attack, an untimed measure of nonword reading, and a mean standard score at the 42nd percentile on Picture Vocabulary, a receptive measure of vocabulary knowledge. In contrast, Cluster 1 earned a mean standard score at the 13th percentile on Sight Word Efficiency, a timed measure of word reading, and a mean score at the 2nd

Table 1 Descriptive statistics of standardized factor scores by cluster

	Mean factor score (SD)			
	Vocabulary factor	Efficiency factor	Decoding factor	Text level factor
Cluster 1	.87 (.58)	.06 (.68)	.52 (.73)	−.26 (.57)
Cluster 2	.38 (.69)	.77 (.63)	.33 (.63)	1.43 (.49)
Cluster 3	−.72 (.48)	.48 (.50)	.20 (.72)	.15 (.56)
Cluster 4	−.42 (.63)	−1.16 (.75)	−1.01 (.70)	−.96 (.56)
Total group	.00 (.88)	.00 (.97)	.00 (.92)	.00 (.98)

Table 2 Marker variable scores for four clusters

	N	Marker variable mean score (SD)			
		SWE	Word attack	Passage accuracy	Picture vocabulary
Cluster 1	38	83.16	100.26	4.47	97.23
Cluster 2	27	91.07	96.33	8.33	91.04
Cluster 3	39	86.92	97.74	5.95	75.10
Cluster 4	36	71.08	86.03	2.81	80.72
Total sample	140	82.63	95.14	5.17	85.64

Note. SWE = Sight Word Efficiency

percentile on passage reading accuracy. This pattern would seem to suggest a specific need for intervention that targets the development of efficiency in support of sublexical, word-level, and text-level skills.

Cluster 2

Cluster 2 contained the highest proportion of third graders (78%), of Black children (20%), and of children categorized as showing no deficit in reading (85%). As reflected in Table 2, Cluster 2 showed a mean standard score at the 39th percentile on Word Attack, an untimed measure of nonword reading. While this score was well within the average range, Cluster 2 earned a mean standard score at the 27th percentile on Sight Word Efficiency (the timed measure of word reading), a mean standard score at the 27th percentile on Picture Vocabulary (the receptive measure of vocabulary knowledge), and a mean score at the 25th percentile on passage reading accuracy. These three marker variable scores are all at the low end of the average range, suggesting areas of vulnerability but not deficit.

Cluster 3

Cluster 3 had a high proportion of second graders (59%), of children exhibiting a single phonological deficit (23%), and the highest proportion of English language learners (75%). Ninety percent of the children in Cluster 3 whose parents provided income information were categorized as low SES. The profile of literacy scores exhibited by Cluster 3 is similar to the profile exhibited by Cluster 1, with a mean untimed decoding score well within the average range (45th percentile), but sight word efficiency and passage accuracy scores in the below average range (at the 19th and 9th percentiles, respectively). However, Cluster 3 also showed a mean receptive vocabulary score at the 5th percentile, suggesting that vocabulary knowledge is a particular weakness in this group. As with Cluster 1, this pattern appears to suggest a specific need for intervention that targets the development of efficiency in support of sublexical, word-level, and text-level skills. In conjunction with this intervention, however, instruction should likely focus on developing students' oral language proficiency in English.

Cluster 4

There were 36 children in Cluster 4, a group with a sizeable proportion of English language learners (66%), low SES children (91%), and children showing a double deficit in phonological processing and rapid naming (25%). Cluster 4 shows a relatively flat profile, with mean scores on Sight Word Efficiency, passage reading accuracy, and picture vocabulary well below average. In contrast with the first three clusters, whose mean untimed decoding score was well within the average range, Cluster 4's mean score on this measure was 86, in the low average range, suggesting that many of these children show difficulties with the accuracy of their phonological decoding skills. Thus this group would likely benefit from systematic instruction in phonics and word identification skills in support of beginning text reading. Additionally, this group would likely require instruction focusing on the development of oral language proficiency in English.

The percentages of subjects in each cluster within various demographic categories are presented in Table 3. A series of chi square analyses was conducted on the demographic categories by cluster membership in order to examine cluster differences in the concentration of children by gender, grade, language, ethnicity, and SES. Results showed that the percentage of participants that were female did not differ by cluster ($\chi^2(3, N = 140) = 3.759, p > .05$). The percentage of participants that were in third grade did differ by cluster ($\chi^2(3, N = 140) = 12.640, p = .005$). As Table 3 shows, Cluster 1 had an unexpectedly high percentage of second graders, while Cluster 2 had an unexpectedly high percentage of third graders. The percentage of participants that spoke a first language other than or in addition to English also differed by cluster ($\chi^2(3, N = 135) = 15.543, p = .001$). As Table 3 shows, Cluster 1 had an unexpectedly low percentage of children who spoke a language other than English as at least one of the first languages in the home. Cluster 3 had an unexpectedly high percentage of children who spoke a language other than English as at least one of the first languages in the home. These results were paralleled by the chi square results on the percentage of Hispanic children in each cluster ($\chi^2(3, N = 134) = 11.331, p = .010$). The comparisons by cluster in the percentage of Black participants and the percentage of low SES participants could

Table 3 Demographics for four clusters

	N	Demographics					
		% Female	% Grade 3	% ELL	% Latino	% Black	% Low SES
Cluster 1	38	29	37	32	41	5	68
Cluster 2	27	52	78	40	52	20	84
Cluster 3	39	36	41	75	76	8	90
Cluster 4	36	42	56	66	69	17	91
Total sample	140	39	51	51	60	12	84

Note. ELL = English Language Learner (as identified by parent report that a language other than English was among the first languages spoken at home); low SES = Low Socioeconomic Status (as identified by parent report that the father's income was less than $35,000 a year) not be evaluated because the chi square assumption of expected values greater than five in at least 80% of the cells was violated.

Research question 3: cluster differences in prevalence of reading disability subtypes

Table 4 shows the percentages of participants in each cluster within reading disability (RD) categories. A series of chi square analyses was conducted on the RD categories by cluster membership in order to examine cluster differences in the concentration of children with no deficit, double deficits, and single deficits in rapid naming and phonological awareness. Results showed that the percentage of participants categorized as showing no deficit differed by cluster ($\chi^2(3, N = 140) = 16.336$, $p = .001$). As Table 4 shows, Cluster 2 had an unexpectedly high percentage of participants categorized as no deficit, while Cluster 4 had an unexpectedly low percentage of participants categorized as no deficit. Similarly, while an examination of Table 4 suggests that Cluster 4 contained an unexpectedly high percentage of participants categorized as double deficit, chi square analyses examining the percentages of double deficit participants by cluster violated the assumption of expected values greater than five in at least 80% of the cells due to the small observed sample size of this RD category. The statistics displayed in Table 4 suggest that Cluster 3 contains an unexpectedly high percentage of participants with a single phonological deficit, although the assumption of expected values was again violated due to the small sample size. An examination of Table 4 also suggests that Cluster 1 contained an unexpectedly high percentage of participants with a single naming speed deficit, although the chi square results testing differences in the prevalence of single naming speed deficit participants by cluster only approached significance ($\chi^2(3, N = 140) = 7.679$, $p = .053$).

Discussion

In response to the persistent achievement gap in reading currently plaguing our nation's urban schools and in particular affecting English language learners, children from low SES families, and minority groups, the *NCLB* Act has focused

Table 4 Categories of reading disability (RD) by cluster

	N	RD Categories			
		% Naming speed deficit only	% Phonological deficit only	% Double deficit	% No deficit
Cluster 1	38	32	3	0	66
Cluster 2	27	7	7	0	85
Cluster 3	39	13	23	3	62
Cluster 4	36	25	14	25	36
Total sample	140	20	12	7	61

national attention on improving instruction for children in failing schools. A necessary means to this end involves the identification of subtypes of children who show unique profiles of strength and weakness in reading skills, allowing teachers to tailor comprehensive reading interventions specific to those profiles. While one previous study (Riddle Buly & Valencia, 2002) has begun this process with upper elementary school children who have failed state reading tests, this study was designed to examine the heterogeneity among a sample of 140 second and third grade urban children identified as at risk for reading failure.

The first research question asked whether the profile suggested by the sample averages on four factors underlying reading performance characterized the entire sample, or whether there were different compelling patterns of performance in the sample. The results of cluster analyses suggest that there are four distinct reading profiles in this sample, none of which parallels the profile suggested by the sample averages. The second research question examined the instructional implications of the identification of these four subgroups of children. The use of marker variables for each of the reading factors allowed a characterization of the reading performance of each subgroup in comparison to national norms (see Table 2). Cluster 1 showed below average scores in sight word efficiency and passage reading accuracy despite average scores in word attack and picture vocabulary. Cluster 2 showed low average scores in sight word efficiency, passage reading accuracy, and picture vocabulary, despite average scores in word attack. Cluster 3 showed below average scores in passage reading accuracy and picture vocabulary, and low average scores in sight word efficiency, despite average scores in word attack. Finally, Cluster 4 showed scores well below average in efficiency and passage reading accuracy, with below average scores in picture vocabulary and low average scores in word attack.

These patterns of performance are clearly related to the literacy instruction these children have already been receiving. All five schools serving the children in this sample use a systematic phonics program as part of the second and third grade language arts curriculum. However, there was considerable variation in the nature of the reading programs used, with some classrooms using whole class basal instruction and other classrooms using Guided Reading with flexible groupings of children. It is likely that the proportion of children in each group, and further that the particular profile exhibited by each group, will vary based on the quality of reading instruction in the sample and the degree to which children are offered instruction appropriate to their needs. Thus rather than offering a picture of universal subtypes of developing readers, these findings should be seen as converging evidence for the utility of assessing children regularly using various measures targeting multiple facets of reading skill and the importance of providing individualized instruction based on assessment results.

Implications for instruction

Despite these caveats, the cluster profiles presented here present interesting questions about the kinds of literacy interventions that would be most suited to children showing these or similar profiles. While this study was not designed to offer definitive conclusions regarding which types of instruction would be most effective for the clusters of children described here, research on effective literacy instruction offers some interesting hypotheses about what instruction might look like for these groups of struggling readers.

Among the children in Cluster 1, which consisted of a high proportion of English monolinguals, vocabulary knowledge and decoding accuracy were relative strengths (likely due in part to the strong phonics instruction these children have received previously). Therefore, reading programs with a heavy emphasis on phonics may not be ideal for this group of children who instead require an intervention focused on their area of difficulty—the development of fluency at the letter, letter pattern, word, text, and meaning levels (Wolf & Katzir-Cohen, 2001). At the sub-lexical level, fluency instruction should emphasize automatic recognition and naming of letters and high frequency letter patterns (Wolf, Miller, & Donnelly, 2000). Word-level instruction should focus on automatic recognition of high frequency and irregular words and fluent application of decoding skills for less common regular words (Rasinski & Padak, 2001). At the text level, the intervention should target the development of accuracy, fluency, and comprehension using carefully monitored instructional level text (Fountas & Pinnell, 1996). With connected text, assisted readings and repeated readings have been shown to promote the development of reading fluency somewhat among beginning readers (Rasinski, 2003; Snow et al., 1998). Nevertheless, it is well documented that gains in fluent comprehension are the hardest to make (Lyon & Moats, 1997). For the children in Cluster 1, fluency instruction would complement their existing strengths in decoding and receptive vocabulary. Improved efficiency at all levels is critical to improving the rate, accuracy and comprehension of text level reading.

In contrast to the children in Cluster 1, who showed marked inefficiency in words level skills and poor passage reading, the children in Cluster 2 showed only *vulnerabilities* in efficiency, passage reading, and vocabulary. For the children in Cluster 2, programs with a heavy emphasis on phonics are no longer appropriate; rather, they would benefit from an intervention focused on the continued development of fluency at the word and text levels, with an emphasis on developing endurance (Wolf & Katzir-Cohen, 2001). In this case, fluency instruction would complement existing strengths in decoding. With increased fluency, reading is likely to become easier and more pleasurable, and the children in Cluster 2 will be more likely to expand their vulnerable receptive vocabularies by reading texts on a variety of topics providing exposure to lower frequency vocabulary. Curriculum with an added emphasis on the development of rich vocabulary knowledge would also be ideal for Cluster 2.

Among Cluster 3, consisting of a high proportion of English language learners, vocabulary knowledge was a marked area of weakness and decoding accuracy was a relative strength. Like Cluster 1, the children in Cluster 3 would benefit from an intervention focused on the development of fluency at the letter, letter pattern, word, text, and meaning levels (Wolf & Katzir-Cohen, 2001; Wolf et al., 2000). However, Cluster 3 would also benefit from integrated instruction in the vocabulary (and likely the syntax) of the English language. Research on effective vocabulary instruction highlights the importance of choosing useful and interesting words for instruction, providing child-friendly definitions, getting children involved in thinking about and using new words, and helping students to maintain their familiarity with new words (Beck, McKeown, & Kucan, 2002).

These results suggest that the children in Cluster 4 bring few strengths to the task of learning to read. In addition to difficulties with automaticity of word and text reading and gaps in English vocabulary comprehension and production, these children show vulnerabilities in phonological decoding accuracy. The causes of reading failure among this group may be heterogeneous. Certainly the majority of these children are just learning English. For children in Cluster 4, the decision must be made whether literacy instruction should be offered in the native language or in English (Snow et al., 1998). If the resources are not available to provide literacy instruction in the child's native language, a systematic program designed to promote oral language proficiency in English should precede an intensive reading intervention in English.

However, 34% of the children in Cluster 4 are native English speakers. Careful consideration must be given to how to remediate their language deficits. Furthermore, these children will require a comprehensive literacy intervention, beginning with instruction targeting phonological awareness, decoding, and word identification and moving toward instruction in text reading and comprehension, with careful monitoring to ensure that progress is being made.

These results also offer support for several general conclusions regarding effective reading instruction for urban children at risk for reading failure, although intervention studies with this population will be necessary for confirmation of these observations. First, systematic instruction in phonics appears to be effective at building decoding skills in these at-risk groups but does not appear to be sufficient to improve their overall reading. The results presented in the previous section show that three out of four of the clusters showed mean decoding scores in the average range. A closer look at the descriptive data for the sample as a whole indicates that 81% of the sample earned a standard score of 90 or above on the Word Attack subtest of the WRMT-R. Moreover, 89% of the sample earned standard scores within one standard deviation of the mean on this measure. Indeed this measure of untimed nonword reading has the highest mean score of any of the literacy measures administered to this sample of struggling readers. Despite their facility with applying phonological decoding strategies to the pronunciation of nonwords, these children still had marked difficulty with tasks requiring efficiency and passage reading. Of particular concern, the children in this sample had difficulty reading nonwords efficiently despite the fact that they could read them accurately in an untimed format. Furthermore, many children had difficulty evaluating the effectiveness of their decoding efforts in the context of passage reading. Many children left miscues uncorrected, even when the miscues violated the meaning of the sentence.

These observations suggest that phonics instruction should go beyond the development of accuracy by focusing on efficiency and the transfer of decoding skills to text reading. Research will need to explore the most effective instructional techniques for achieving these goals, but in the meantime, some promising instructional techniques have been identified. To promote the development of efficiency in decoding skills, instruction should include methods that have proven to be effective at promoting efficiency in reading more generally, including assisted readings and repeated readings (in this case of words or text containing the target letter patterns) (National Reading Panel, 2000). There is also some support for the use of timed readings to promote the development of fluency (Sample, 2005), suggesting that timed exercises with words grouped by letter pattern might promote the development of efficient decoding skills (e.g., Great Leaps fluency program). With regard to the transfer of decoding skills to text reading, research has shown that one-on-one tutoring using systematic word study in combination with text reading practice resulted in greater gains in passage reading fluency and accuracy among first grade poor readers than systematic word study alone (Vadasy, Sanders, & Peyton, 2005). Consistent with this finding, Biemiller (1994) argued that it is not enough to provide children with new skills and competencies; we must provide them with opportunities to apply those competencies for true development to occur. In the present case, it seems clear that struggling readers need more than systematic phonics instruction, even with an increased emphasis on the development of efficiency with these skills; they require scaffolding as they apply their newfound decoding skills to an authentic task, the enjoyment of a good book. Effective reading instruction will need to offer opportunities for both.

Second, comprehensive reading instruction should include a focus on the development of efficiency at multiple levels, based on a multi-componential model of reading fluency (Berninger, Abbott, Billingsley, & Nagy, 2001; Wolf & Katzir-Cohen, 2001; Wood, Flowers, & Grigorenko, 2001). Wolf and Katzir-Cohen (2001) advocated that ''systematic instruction should be directed to accuracy and then to rate at each developmental level of the acquisition of reading subskills—that is, at the level of the phoneme, grapheme, letter, letter pattern (orthographic chunk), word, phrase and sentence, and passage'' (p. 229). One intervention designed to address this set of levels is the RAVE-O program (Retrieval, Automaticity, Vocabulary, Engagement, and Orthography), designed to provide a systematic, comprehensive fluency instruction for children exhibiting deficits in rate of processing and reading fluency (Wolf et al., 2000). RAVE-O emphasizes rapid recognition of frequent orthographic letter patterns, alongside vocabulary development, increased lexical retrieval and syntactic skills, and improved passage reading rate. These results suggest that several of the subgroups of children studied here would benefit from more tailored instruction focusing on the development of efficiency in multiple, specific domains.

Finally, given the fact that only 39% of the children in our sample attained receptive vocabulary scores in the average range (at or above a standard score of 90), many of the children in this population of at-risk readers need a well-coordinated language and literacy program, with instruction designed to promote the development of reading skills and the development of language production and comprehension. For children who enter school with little or no proficiency in English, the instructional priority should be to develop oral proficiency in English with reading instruction if feasible in their native language (Snow et al., 1998). But low levels of language proficiency affect academic performance among native English speakers as well, due to vocabulary and syntax differences associated with

SES and dialect as well as individual variation in verbal ability (Hart & Risley, 1995; Washington & Craig, 1999). Therefore, a wide range of students would benefit from integrated reading instruction within a classroom environment steeped in rich language, including exposure to high quality literature (Morrow, 1992; Robbins & Ehri, 1994), exposure to different genres of text (Caswell & Duke, 1998; Duke, 2004), opportunities to write for authentic purposes (Graham & Harris, 1997), and robust vocabulary instruction focusing on depth and breadth of word knowledge (Beck et al., 2002).

Categories of RD among urban poor readers

The third research question examined the prevalence of various categories of RD among these four clusters of urban dysfluent readers. These results should be replicated using larger samples and additional measures allowing for more reliable categorization into the categories of RD. Nevertheless, the trends in this first look suggest that each cluster was more closely associated with one category of RD. Cluster 1 contained the highest percentage of children categorized as showing a naming speed deficit, and Cluster 3 contained the highest percentage of children categorized as showing a phonological deficit. Cluster 2 consisted primarily of children exhibiting no deficit, while Cluster 4 included all but one of the children in the sample exhibiting a double deficit. The results of this study offer converging evidence that children showing a double deficit are in fact more disabled than children showing a single deficit in either rapid naming or phonological processing (Lovett, Steinbach, & Frijters, 2000; Manis, Doi, & Bhadha, 2000). Additionally, children showing no deficit in either phonological processing or rapid naming appear to be the least disabled along the four factors underlying literacy performance.

It is important to reiterate, however, that this sample was largely comprised of children who did not meet criteria for a diagnosis of reading disability. Each cluster consisted of a smaller group of children whose phonological processing or rapid naming ability undermined their reading development, and a larger group of children with similar profiles along the four factors but whose rapid naming and phonological processing were in the average to above average range. While further research is necessary to fully explain who these groups of nonreading-disabled urban children are and why they show reading profiles similar to various subtypes of reading disabled children, these results do allow several interesting observations.

Cluster 1 contained both the highest percentage of children exhibiting a single naming speed deficit and the highest percentage of second graders. This observation might be explained in two ways. One possibility is that the children exhibiting rapid naming deficits were also more likely to be second graders. This explanation suggests that the use of national norms to evaluate the raw scores of urban second graders on rapid letter naming tasks results in the over-identification of children with true rapid naming deficits. While urban children may be slower, on average, to automatize their skills in rapid letter naming, this delay is no longer apparent in third grade, at which point national norms become an appropriate reference. A second possibility, however, is that the children with rapid naming deficits in

Cluster 1 come in equal percentages from second and third grade, but the remaining children in Cluster 1 are more likely to be second graders. This explanation suggests that second graders (without a deficit in rapid naming) are more likely than third graders (without a deficit in rapid naming) to show the Cluster 1 profile—deficits in efficiency of word-level skills and accuracy of text reading. On a more encouraging note, however, this explanation suggests that, by third grade, these deficits in word and text level skills have been ameliorated. This explanation is consistent with the observation, noted above, that Cluster 2 (which shows only vulnerabilities in word reading efficiency and text reading accuracy) contains the highest percentage of third graders. While longitudinal data are necessary to settle this point conclusively, additional analyses on these data offer some insight. First, an analysis of variance indicated that there was no significant difference between second and third graders' mean RAN Letters standard scores. Second, a chi square analysis comparing the percentage of rapid naming only deficit children by grade indicated that the observed trends (i.e., more second grade subjects than third grade subjects showed a rapid naming deficit, and more third grade subjects than second grade subjects showed no rapid naming deficit) were not significant. Thus it seems possible that urban second graders are more likely than urban third graders to show a reading profile similar to reading disabled children exhibiting a single deficit in rapid naming ability (though other explanations, for example a cohort effect, are equally likely, underscoring the need for longitudinal studies of this issue).

Another interesting observation comes from a closer examination of Cluster 3, a group of children containing the highest percentage of ELLs and children showing a single phonological deficit. Despite earning low scores on the phonological elision task (mean score = 7.6), the children in Cluster 3 earned average scores on the nonword reading task (mean score = 97.7). The difference in these mean scores is surprising given previous research on the connection between phonological awareness and decoding. Phonological awareness tasks have repeatedly been shown to predict decoding skill, particularly in the early grades (Juel, 1988; Liberman, Shankweiler, & Liberman, 1989, and see Snow et al., 1998 for a review). Some researchers have suggested that phonological awareness and decoding have a reciprocal relationship whereby phonological awareness allows for the development of beginning decoding skill and the development of decoding skill allows for a more sophisticated appreciation of the phonological structure of words (Ehri, 1987; Hogan, Catts, & Little, 2005; Perfetti, Beck, Bell, & Hughes, 1987; Torgesen et al., 1994).

So one explanation of these findings is that the children in Cluster 3 developed sufficient phonological awareness to support the development of decoding skills, but their decoding knowledge failed to promote a more sophisticated awareness of the phonological structure of words, which would have supported performance on the phonological elision task. Another explanation, however, is that phonological elision task was an impure measure of phonological awareness in this sample of ELLs due to measurement error associated with comprehension of directions for this complex task. Further analyses offer little support for this second explanation. The scores on the Elision test were not significantly correlated with receptive vocabulary as measured by the PPVT-III. Moreover, results showed that the percentage of participants exhibiting phonological deficits did not differ by category of first language learned. Clearly, future research will need to provide longitudinal data on the relationship between the developments of phonological awareness and decoding among urban children many of whom are ELLs.

Conclusion

Classroom teachers in urban elementary schools will hardly be surprised by the findings of this study. One of the greatest challenges of teaching is accommodating the various instructional needs of a heterogeneous grouping of children. However, recent changes in educational policy moving toward standardization of reading instruction, however well intentioned, have made it next to impossible to meet individual needs in urban classrooms. This study reminds us of the importance of skilled teachers (and teacher training), of regular assessment of student progress using a variety of measures, and of individualized instruction based on learning profiles. Focusing on these multiple elements of effective instruction is critical to ensuring that no child is left behind.

Acknowledgements This study was part of a reading and social emotional intervention project (The New 3R's—Reading, Resilience, and Relationships in Afterschool Programs) conducted at Harvard University, Tufts University, and McLean Hospital, and funded by the Department of Education, awarded under the Interagency Education Research Initiative (IERI) (CFDA #84.305W), and the Piper Foundation, USA. The authors would like to thank Sally Wilson for her work coordinating this study, and Adrianna Wechsler, Larina Mehta, and Chessie Shaw for assistance with data collection and entry. We are also grateful to Nonie K. Lesaux, Terry Tivnan, Catherine Snow, and three anonymous reviewers, who provided important feedback in improving the write-up of the study.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research review* (pp. 71–117). Newark, Del.: International Reading Association.

August, D., & Hakuta, K. (Eds.) (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Academy Press.

Beck, I., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary development*. New York: Guilford Press.

Berninger, V. W., Abbott, R. D., Billingsley, F., & Nagy, W. (2001). Processes underlying timing and fluency of reading: Efficiency, automaticity, coordination, and morphological awareness. In M. Wolf (Ed.), *Time, fluency, and dyslexia*. Timonium, MD: York Press.

Biemiller, A. (1994). Some observations on beginning reading instruction. *Educational Psychologist, 29*, 203–209.

Caswell, L. J., & Duke, N. K. (1998). Non-narratives as a catalyst for literacy development. *Language Arts, 75*, 108–117.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African-American children and its relation to early reading achievement. *Child Development, 75*(5), 1340–1356.

Duke, N. K. (2004). The case for informational text. *Educational Leadership, 61*(6), 40–44.

Duncan, G. L., & Seymour, P. H., (2000). Socioeconomic differences in foundation level literacy. *British Journal of Psychology, 91*, 145–166.

Dunn, L., & Dunn, L. (1997). *Peabody picture vocabulary test III*. Circle Pines, MN: American Guidance Service.

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*, 5–31.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York, NY: John Wiley and Sons.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10. .

Graham, S., & Harris, K. R. (1997). It can be taught, but it does not develop naturally: Myths and realities in writing instruction. *School Psychology Review, 26*, 414–424.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Co.

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools, 36*, 285–293.

Huisingh, R., Barrett, M., Zachman, L., Blagden, C., & Orman, J. (1990). *The elementary word-r test: A test of expressive vocabulary and semantics*. Lingui Systems, Inc.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology*, *80*, 437–447.

Kameenui, E., Carnine, D., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly, 17*(3), 367–388.

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the gray oral reading test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10*(4), 363–380.

Labov, W. (1995). Can reading failure be reversed: A linguistic approach. In V. Gadsden & D. Wagner (Eds.), *Literacy among African-American youth: Issues in learning, teaching, and schooling*. Cresskill, NJ: Hampton.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 30*(1), 159–174.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–31). Ann Arbor, MI: University of Michigan Press.

Lovett, M. W., Steinbach, K. A., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double deficit perspective. *Journal of Learning Disabilities, 33*(4), 334–358.

Lyon, G. R., & Moats, L. C. (1997). Critical conceptual and mehodological considerations in reading intervention research. *Journal of Learning Disabilities, 6*(30), 578–588.

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33*, 325–333.

Mehta, P. D., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005) Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1–4. *Scientific Studies of Reading, 9*(2), 85–116.

Morrow, L. M. (1992). The impact of a literature-based program on literary achievement, use of literature, and attitudes of children from minority backgrounds. *Reading Research Quarterly, 27*, 250–275.

National Reading Panel (2000). *Report of the national reading panel: Teaching children to read* (NIH Pub. No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.

No Child Left Behind Act of 2001 (2002). Pub. L. No. 107-110. 115 Stat. 1425.

Noam, G., Wolf, M., & Katzir, T. (2003). *The new 3R's—Reading, resilience, and relationships in afterschool programs: A systematic research intervention to increase academic performance.* Grant proposal submitted to Interagency Education Research Initiative.

Norusis, M. (2006). *SPSS 14.0 statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.

Perfetti, C. A., Beck, I. L., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283–319.

Pierce, M., Katzir, T., Wolf, M., & Noam, G. (2006). *Examining the heterogeneity in literacy profiles among urban struggling readers*. Paper presented at the annual meeting of the Society for the Scientific Study of Reading, Vancouver, BC.

Raskinski, T. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency and comprehension*. New York, NY: Scholastic.

Rasinski, T. V., & Padak, N. D. (2001). *From phonics to fluency: Effective teaching of decoding and reading fluency in the elementary school*. New York: Addison Wesley Longman. .

Riddle Buly, M., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading tests. *Educational Evaluation and Policy Analysis, 24*, 219–239.

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology, 86*(1), 54–64.

Sample, K. (2005). Promoting fluency in adolescents with reading difficulties. *Intervention in School & Clinic, 40*(4), 243–246.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276–286.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level identification in supplemental tutoring: How text matters. *Journal of Learning Disabilities, 38*, 364–380.

Valencia, S. W., & Riddle Buly, M. R. (2004). What struggling readers REALLY need. *Australian Journal of Language and Literacy, 27*, 217–233 (reprinted from *The Reading Teacher, 57*, 520–533).

Wagner, R., Torgesen, J., & Rashotte, C. (1999). *Comprehensive test of phonological processing*. MN: American Guidance Service.

Washington, J., & Craig, H. (1999). Performances of at-risk, African American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools, 30*, 75–82.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray oral reading tests* (4th edn.). Austin, TX: Pro-Ed.

Wolf, M., & Denckla, M. B. (2004). *Rapid automatized naming and rapid alternating stimulus tests*. Austin, TX: Pro-Ed.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*, 211–239.

Wolf, M., Miller, L., & Donnelly, K. (2000). The retrieval, automaticity, vocabulary elaboration, orthography (RAVE-O): A comprehensive fluency-based reading intervention program. *Journal of Learning Disabilities, 33*(4), 375–386.

Wood F. B., Flowers D. L., & Grigorenko E. (2001). The functional neuroanatomy of fluency or why walking is just as important to reading as talking is. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain*. Baltimore, MD: York Press.

Woodcock, R. (1997). *Woodcock reading mastery test-revised*. Circle Pines, MN: American Guidance Service.