

1287

Late-Emerging Reading Disabilities

Jennifer Mirak Leach
Bryn Mawr College and Haskins Laboratories

Hollis S. Scarborough
Haskins Laboratories

Leslie Rescorla
Bryn Mawr College

Literacy, language, and cognitive skills were compared for 35 4th–5th graders with early-identified reading disabilities (RD), 31 with late-identified RD (first seen after 3rd grade), and 95 normally achieving students. Late-identified reading deficits were heterogeneous; some children were weak in both comprehension and word-level processing, whereas others had deficiencies in 1 component of reading but were unimpaired in the other. Although most reading skill deficits were about as severe for late- as for early-identified RD, and profiles of associated characteristics were similar, few of the former had yet been identified by their schools. Third-grade achievement, retrospectively examined, had been higher for the group with late-identified RD, suggesting that their reading difficulties were not just late identified but actually late emerging.

A great deal of knowledge and understanding about children's reading disabilities has been amassed over the past few decades. The focus of most research has been early-emerging difficulties that are identified prior to the fourth grade. As noted by Chall (1983), however, there also occurs a smaller "second wave" of students who appear to undergo a "fourth-grade slump" in reading achievement. There has been almost no investigation of the nature and severity of these less typical reading disabilities (RD) that emerge after the primary grades. In this study, we compared the reading and spelling abilities and literacy-related skills of children with early- versus late-emerging RD.

In practice, during the elementary and middle school years, there is typically an increasing emphasis on reading comprehension in curricula, instruction, and assessments (Anderson, Hiebert,

Scott, & Wilkinson, 1985; Applebee, Langer, & Mullis, 1987). As Chall (1983) aptly put it, the main focus during the primary grades is "learning to read," but thereafter it becomes "reading to learn." Even within contemporary meaning-emphasis curricula, beginning students are still expected to grasp the alphabetic principle (that the phonological elements of spoken language are represented by the orthography of English), to become familiar with the systematic correspondences between letters and phonemes (and with exceptions to these regularities), and to gain facility in applying this knowledge (i.e., in decoding and encoding print) so as to become quite proficient at word recognition and spelling by the third grade. Most texts that these children read or listen to are narratives that pose relatively little challenge with regard to vocabulary and comprehension, particularly if the material is written so as to be within the current word recognition capabilities of beginning readers (Chall, Jacobs, & Baldwin, 1990; Duke, 2000; Hayes & Ahrens, 1988).

Similarly, both formal and informal assessments of reading in the early grades—from "running records" to standardized achievement batteries—typically include some evaluation of the student's speed and accuracy in recognizing printed words. Children who are not progressing satisfactorily in acquiring word recognition skills are likely to be referred for a fuller evaluation of the possible need for special assistance. It is well documented that a majority of young children with RD are classified as such on the basis of weak word-level processing skills rather than comprehension deficits alone (Nation & Snowling, 1997; Shankweiler et al., 1999; Yuill & Oakhill, 1991) and that their lags in acquisition are sufficiently evident by second or third grade to merit a formal referral by the classroom teacher or other school personnel (Kavale & Reese, 1992).

Beyond the primary grades, less instruction in word recognition and spelling skills is offered, and the emphasis shifts to honing reading comprehension skills and strategies. Written materials in

Jennifer Mirak Leach, Department of Psychology, Bryn Mawr College, and Haskins Laboratories, New Haven, Connecticut; Hollis S. Scarborough, Haskins Laboratories; Leslie Rescorla, Department of Psychology, Bryn Mawr College.

This report is based on a dissertation by Jennifer Mirak Leach in partial fulfillment of the requirements for a doctoral degree in psychology at Bryn Mawr College. The research was supported, in part, by Program Project Grant HD01994 to Haskins Laboratories. Preliminary reports of the results were presented at conferences of the Society for the Scientific Study of Reading, Montreal, Quebec, Canada, April 1999, and the International Dyslexia Association, Washington, DC, November 2000.

We thank the children and their parents for their participation and the many administrators and teachers at the children's schools for their cooperation and encouragement. We thank Jarlette Lampard for assistance in collecting the data and our colleagues Kimberly W. Cassidy, Carol Roberts, and Barbara J. Lorry for their many helpful comments on all aspects of the project.

Correspondence concerning this article should be addressed to Hollis S. Scarborough, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511. E-mail hscarborough@prodigy.net

language arts and in other academic subjects become increasingly challenging with regard to their length, vocabulary level, syntactic complexity, and conceptual demands. Similarly, text comprehension becomes the main focus of both low-stakes and high-stakes assessments, although reading vocabulary tests continue to be administered in many schools. Given these shifts over time in what is taught and what is assessed, poor performance on reading comprehension tests in fourth grade and beyond is often the first indication that a student may have fallen behind after a successful start in reading acquisition.

It is generally presumed, therefore, that the basis for the child's late-emerging difficulties is a weakness in higher order comprehension skills. Indeed, strong word-level skills in conjunction with weak comprehension is a profile that has been observed in older children and in low-literate adults (Spear-Swerling, 2001; Strucker, 1995). There is considerable evidence that difficulties in reading comprehension are likely to be accompanied by deficiencies in oral language (Nation & Snowling, 1998; Stothard & Hulme, 1992), background knowledge (Chi & Koeske, 1983; Pearson, Hansen, & Gordon, 1979), metacognitive awareness (Anderson, 1980; Wong & Wong, 1986), strategy use (Hare & Pulliam, 1980; Kletzein, 1991), and/or memory capacity (Just & Carpenter, 1992; Swanson, Cochran, & Ewers, 1990). Consequently, schools generally provide special assistance designed to strengthen the comprehension-related skills and strategies of postprimary students with RD.

Not all cases of late-identified reading difficulties may involve deficits in comprehension, however. Beyond the primary grades, heavier demands are made not just for deeper understanding of text but also for dealing with more challenging aspects of lower-level processes. For example, decoding and spelling of complex polysyllabic words, and familiarity with orthographic patterns associated with Anglo-Saxon, Greek, and Latin root morphemes, are expected to be mastered once the rudimentary phoneme-grapheme correspondences have been learned. Moreover, bottom-up processing skills must become increasingly automatized in the later elementary grades, leading to improvements in the speed of word recognition and the fluency of text reading.

It is possible, therefore, that some children's progress in reading acquisition falters when they are faced with these new challenges. Indeed, weaknesses in bottom-up processing have been observed to persist even in college students and adults with a childhood history of RD (Bruck, 1990; Scarborough, 1984; Wilson & Lesaux, 2001) and to be seen in a substantial proportion of adults with low literacy skills (Fowler & Scarborough, 1993; Strucker, 1995). In some cases, therefore, RD might be identified in fourth grade or later if a mild or well-disguised reading difficulty in the primary grades becomes more severe and debilitating in response to increasing demands for greater accuracy and especially speed of decoding. As Juel (1991) and others have noted, some children rely almost entirely on sight memorization of words and thereby appear to be succeeding in learning to read for the first several years until this strategy becomes unwieldy. Even children who initially master the rudiments of decoding successfully, however, could nevertheless encounter difficulty when the words to be recognized and spelled become more phonologically and morphologically complex after the primary grades.

Moreover, if bottom-up skills are inaccurate or inefficient, comprehension is also likely to suffer. That is, when recognizing the words in connected texts is slow and effortful, fewer cognitive resources can be devoted to the higher level processes needed to attain an understanding of the text's meaning (LaBerge & Samuels, 1974; Perfetti, 1985). Hence, it is argued, low reading comprehension scores at older ages might often be attributable to difficulties with word-level processing rather than to true deficits in understanding. In some cases, of course, a late-emerging RD could involve both top-down and bottom-up difficulties, a profile that has been observed quite frequently in older children and low-literate adults (Spear-Swerling, 2001; Strucker, 1995).

In sum, a late-emerging RD might involve difficulties in text comprehension, in word-level processing, or in both. Children with deficits only in comprehension would be expected to read and spell words and pseudowords quickly and accurately but to exhibit marked weaknesses in both reading and listening comprehension. Vocabulary skills might also be impaired, but phonemic awareness and other phonological skills should be intact. Children with deficits only in word-level processing would be expected to be inaccurate and/or slow at reading and spelling words and pseudowords and to have associated weaknesses in phonemic awareness and symbol naming speed. Poor performance on reading comprehension measures would also be likely because the hypothesized word recognition "bottleneck" would divert cognitive resources from processes necessary for obtaining meaning from connected text. Listening comprehension, however, should be unimpaired unless the child has deficits in both top-down and bottom-up aspects of reading.

Although there are indications that late-identified cases constitute a nontrivial proportion of all students with RD (Badian, 1999; Kavale & Reese, 1992; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992), there has been almost no formal research on the characteristics of these children. Our goal was to examine the cognitive and achievement profiles of this understudied population and to compare them with those of early-identified disabled readers. Three main questions were addressed: How heterogeneous are children with late-identified reading disabilities with regard to their strengths and weaknesses in various components of reading skill? In comparison with early-identified cases, do late-identified disabled readers have less severe, less broad, or altogether different profiles of abilities? Does late identification of the reading problem occur because schools have overlooked earlier weaknesses, or because the child's reading difficulties actually emerge later?

Method

Participants

The 161 participants were 74 fourth graders (35 boys and 39 girls) and 87 fifth graders (48 boys and 39 girls) from 12 elementary schools in two neighboring school districts in the greater Philadelphia area. Six of the schools served more affluent neighborhoods in which 2.5% to 8.8% of students qualified for the federal lunch program. The other six schools were more socioeconomically diverse, with 12.4% to 59.6% of students receiv-

ing free or reduced-cost lunches.¹ Only 5% of the children in the sample were ethnic minorities (African American, Asian, or Hispanic), the others being Caucasian. Ages ranged from 9.1 to 12.0 years ($M = 10.5$, $SD = 0.7$).

The participants were selected on the basis of information provided voluntarily by parents. All fourth- and fifth-grade students in the two districts (approximately 2,300 children) were given a packet of materials by their classroom teachers and asked to take them home. Each packet contained a letter explaining the study; a parent questionnaire; a parent consent form; and a stamped, addressed, return envelope. The study was described as an investigation of reading skills in which students of all reading-ability levels were eligible to participate.

Completed questionnaires and signed parental consent forms were returned for 385 students. Of these, the pool of eligible participants was narrowed to 289 children who met four criteria: (a) they were native speakers of English; (b) they first enrolled in the district prior to or during the second grade; (c) their parents consented to future phone contact by the researchers; and (d) in those cases where previous test results were in their school files, they had a Full Scale IQ (FSIQ) score above 70. No child failed to meet the last criterion.

Parents also completed a nine-item questionnaire about the child's history of reading problems and any special assistance that was received either in or outside of school. These parental reports of the child's educational history were checked against the school records of each child's history of interventions (Title 1, special education services, etc.), if any, and there was complete agreement about whether a reading difficulty had been identified by the school. A few discrepancies were seen with regard to when such identification took place, and in these instances we relied on the school file. On the basis of the history of reading difficulties derived from parental reports and school records, the 289 children were assigned to one of the following five groups:

Early school-identified: Persistent (ESI-P). These 89 students had been identified by the school as having a reading problem prior to the fourth grade and continued to receive reading-based remediation within 1 school year prior to data collection. A child was considered to have a school-identified reading problem if one or more of the following occurred: a referral, evaluation, or identification as requiring special education services; inclusion in the Title 1 reading program; extra help provided during or after school; concerns communicated to parents by school personnel about unsatisfactory progress in reading.

Early school-identified: Transient (ESI-T). These 22 students had school-identified reading problems (as defined previously), but remedial services (in or outside of school) were discontinued prior to the third grade and never resumed.

Late school identified (LSI). These 13 students had reading difficulties that had been first referred for evaluation in school during fourth or fifth grade.

Parent concern (PAR). These 13 students had never been identified by the school as having a reading problem, but their parents had expressed moderate or strong concern about their child's reading at some point during the primary grades.

No history (NOH). This group included 152 students who had no history of reading difficulties and whose parents had never had more than mild concern about their reading skills.

Given the limited time and resources available, all 289 eligible students could not be tested. Therefore, it was decided to select all students in the LSI and PAR groups, but only about 50% of each other group. Children in the ESI-P, ESI-T, and NOH groups were first sorted by grade and gender, and individuals were then selected randomly from those lists until the target sample size was reached or time ran out, so that roughly equal proportions of fourth and fifth graders and boys and girls per group would be obtained. By the end of the school year, when data collection ceased, 161 children had been tested, distributed as follows: 42 ESI-P, 10 ESI-T, 13 LSI, 13 PAR, and 83 NOH. Across these five groups, there were

slight differences in the proportions of boys (.52, .30, .46, .54, and .48, respectively), of fourth graders (.50, .40, .31, .31, and .49, respectively), and of students from lower socioeconomic status (SES) schools (.50, .50, .69, .54, and .48, respectively), but educational history was not significantly associated with gender, $\chi^2(4, N = 161) = 1.80, p = .77, \phi = .11$; with grade, $\chi^2(4, N = 161) = 3.23, p = .52, \phi = .14$; or with the SES level of the school, $\chi^2(4, N = 161) = 2.05, p = .73, \phi = .12$.

Measures

Information about past and current reading abilities and reading-related skills and attitudes was obtained from individual assessments of the children and from school records.

Child testing. Tests of reading and spelling skills, measures of cognitive and language abilities, and two brief questionnaires were individually administered to each participant. The Word Identification and Word Attack subtests of the Woodcock-Johnson Psychoeducational Battery—Revised (WJ-R; Woodcock & Johnson, 1989) require the student to read aloud lists, of increasing difficulty, of isolated words and pseudowords, respectively. Raw scores (total number correct) were converted to standard scores based on national norms. For the 308 9-year-olds in the standardization sample, there was strong internal consistency (split-half reliability) for both the Word Identification (.94) and Word Attack (.91) subtests (Woodcock & Johnson, 1989).

The Reading Comprehension subtest of the Peabody Individual Achievement Test—Revised (PIAT-R; Dunn & Marquardt, 1989) was used to assess comprehension of both written text and spoken text. Following Spring and French (1990), half of the 76 items were administered in the standardized format, whereas the other half (randomly selected) were administered orally as a listening comprehension measure. This permitted the comprehension of written and spoken items to be compared for material that was equivalent in length and linguistic complexity. For reading, the participants read each sentence silently and then chose one of four pictures that corresponded best to the sentence. Estimated standard scores were derived from norms tables after doubling the raw number of correct responses. For the listening portion, which was administered later in the session, the examiner read the sentence aloud with moderate expression and then presented the four picture choices. For the purpose of comparing performance in the reading and listening formats, standard scores were estimated using the same procedure as for reading. Because this modified procedure was not used in the standardization of the PIAT-R, the standard scores we derived cannot be interpreted in the conventional manner, especially for listening comprehension. Strong internal consistency reliability was found for both reading comprehension scores ($\alpha = .89$) and listening comprehension scores ($\alpha = .87$) in the sample.

Two experimental spelling-to-dictation tests were given. On each, the item was pronounced, used in a sentence, and then pronounced again by the examiner. For Exception Word Spelling, the child was asked to write down 12 words that require word-specific knowledge (i.e., are not entirely regular), chosen because pilot testing suggested that many are not mastered by the fifth grade. The words were as follows: *once, sugar, wonder, natural, doesn't, marriage, unnecessary, building, rhythm, somebody, Pennsylvania, and hieroglyphics*. The proportion correct was scored. For Pseudoword Spelling, the child was asked to spell made-up words as if they were real words. The 10 items were as follows: *shems, druffy, jicks, glessed, zoyable, norger, cathective, venches, qualistic, and exidotionally*. Spellings of each pseudoword were scored both phonologically (to evalu-

¹ Because the association between SES and reading achievement is largely a between-schools (mean $r = .68$) rather than a between-students ($r = .21$) effect (Committee on the Prevention of Reading Difficulties in Young Children, 1998; White, 1982), information about the SES of the individual participants' families was not collected.

ate knowledge of phoneme-grapheme correspondences) and orthographically (to assess understanding of 23 basic orthographic conventions for spelling certain phonemes in particular contexts, and for representing familiar inflectional and derivational morphemes). For example, on the item "glessed (/gles/) . . . I have glessed it three times, but I need to gless it again . . . glessed," a participant who wrote *gless* would be scored as phonologically accurate, because each spoken phoneme is represented by a grapheme that can represent that phoneme in English orthography. The response would not, however, be scored as orthographically accurate with regard to either of two applicable spelling conventions: (a) the *-ed* past tense affix and (b) because the /gles/ portion should be considered the root word, the need to double the final *s*. Internal consistency (α) was .70 for exception word spelling, .82 for phonological spelling of pseudowords, and .72 for orthographic spelling of pseudowords in the sample.

Rapid Serial Naming speed was measured in the conventional manner for three stimulus arrays: letters (*j, r, p, m, h*), digits (6, 4, 2, 5, 9), and pictured objects (*truck, chair, horse, clock, spoon*). Each array of 5 rows of 10 items was displayed on a 25 × 35 cm poster board, with the 5 items each repeated 10 times. Students were asked to name all 50 items per array (in top-down, left-to-right order) as quickly as possible without making errors but to keep going if an error was made. A practice array (of colors) was used to instruct participants. Also, to ensure that the items in the timed arrays were known, the 5 stimulus items for each array were introduced and named beforehand. For each array, the number of seconds taken to name the 50 items was measured and converted to an items-per-second score. Naming rates for letters and digits were highly correlated ($r = .80$), indicating good reliability of symbol naming speed. No estimate was obtained for the reliability of object naming rates, which were not as strongly correlated with the speed of naming letters ($r = .61$) or digits ($r = .62$).

Speed of reading was measured and scored analogously to naming speed. Immediately after the rapid naming arrays were completed, three additional 5 × 10 arrays were presented, each containing 25 three- and four-letter items (each appearing twice in the array). Items in the three arrays were as follows: regularly spelled nouns (e.g., *key, fork, and man*), irregularly spelled high-frequency words (e.g., *was, off, and some*), and pseudowords that were created by changing one letter of each word in the noun array (e.g., *koy, bork, and mab*). The Word Reading Speed score was the average words per second on the two word arrays; Pseudoword Reading Speed was based on one array only. Rates on all three arrays were highly correlated ($r = .84$ to $.85$), indicating that reading speeds were reliably measured.

Our Phonological Awareness measure was similar to the Lindamood Auditory Conceptualization Test (Calfee, Lindamood, & Lindamood, 1973; Lindamood & Lindamood, 1979) but was shorter and contained a higher proportion of items requiring fine-grained phonemic segmentation. Administration began with the examiner explaining that a row of one or more differently shaped tokens would represent the sounds in a syllable. Syllables contained one or more phonemes from a restricted set: the vowel /u/ (*oo*) and the consonants /k/, /l/, and /s/. On two demonstration items and 14 scored trials, the examiner changed the spoken stimulus by substituting, adding, subtracting, or reordering a phoneme and asked the child to modify the token array appropriately. For example, on being told, "If this is /uks/, show me /usk/," the child would have to reverse the positions of the third and fourth tokens in the row. On the fifth and ninth items, the examiner also changed the stimulus by introducing a new token representing another phoneme. The sequence of transformations was as follows: /ku/-/u/; /u/-/uk/; /uk/-/ku/; /ku/-/kuk/; /luk/-/lu/; /lu/-/kul/; /kul/-/kulu/; /kulu/-/klul/; /klus/-/kluks/; /klus/-/lusk/; /lusk/-/skul/; /skul/-/sklu/; /sklu/-/sluk/. Proportion correct was scored, and good internal consistency ($\alpha = .90$) was obtained for these scores.

The scaled score for the Block Design and Vocabulary subtests from the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991) served as measures of nonverbal and verbal cognitive

ability, respectively. From these, an estimate of FSIQ was calculated following the procedures developed by Sattler (1992).

A test of visual-verbal learning was created to simulate "sight learning" of whole words as visual patterns. Participants were told that they would learn some words from a foreign language. A series of six pseudologographs was presented at a rate of 4 s per item, with a spoken English "translation" given for each (*bird, book, chair, monkey, river, tree*). Students were then asked to name the stimuli as they were presented (in different orders) on three recall trials with corrective feedback. The total proportion correct over all 18 trials was scored. Because of ceiling effects, variability of these scores was quite limited, and reliability was somewhat lower than for other measures ($\alpha = .67$).

Questionnaires. To estimate print exposure (i.e., how much a person has read), an age-appropriate title recognition checklist was used. A list of 48 titles was generated by combining items from the Title Recognition Test (Cunningham & Stanovich, 1990) and the Title Recognition Test—Primary (McDowell, Schumm, & Vaughn, 1992). Sixteen actual titles and eight foils were taken from each source. Participants were told that some of the titles were names of actual books and some were not. They were asked to read the list and to check off those they knew for certain were actual books. Guessing was strongly discouraged. True positive and false negative response rates were used to compute an A' score that adjusted for guessing biases (Snodgrass & Corwin, 1988). In addition, self-efficacy ratings were made by the participants of their own reading skills, reading speed, math skills, spelling skills, and overall intelligence in comparison with those of "other kids your age." Ratings were based on a 7-point Likert scale, ranging from 1 (*much worse*) to 7 (*much better*).

School records. From school files, we obtained the children's national percentile scores on nationally standardized group-administered tests of reading vocabulary, reading comprehension, spelling, and math achievement that had been given in third grade in order to document achievement levels at a younger age (i.e., 1 or 2 years prior to our assessments). One district used the Comprehensive Test of Basic Skills (CTBS) and the other used the Iowa Test of Basic Skills (ITBS). No subsequent scores on school-administered tests could be obtained for the sample because neither district gave these tests in fourth grade, and scores were not yet available for the fifth graders.

Missing Data

School records could not be located for 5 children, 3 from the NOH group, 1 from the LSI group, and 1 from the ESI-T group. Hence, third-grade achievement scores were missing for these participants. In addition, the WISC-III measures were not administered to 1 child from the ESI-P group so as not to jeopardize the validity of a clinical evaluation that was scheduled for the following week.

Procedure

Each child was individually assessed in April, May, or June by one of two female examiners who were randomly assigned to participants and were blind to the educational histories of the children. Testing took place during school hours in private testing spaces on the school premises. Each child was seen for a single 1-hr session, which was audio recorded. All children readily assented after being fully informed of the goals and procedures of the research. The task order was as follows: rapid serial naming, reading speed (words, pseudowords), Word Identification, Word Attack, visual-verbal learning, Reading Comprehension, phonological awareness, listening comprehension, exception word spelling, pseudoword spelling, print exposure checklist, Block Design, self-efficacy ratings, and Vocabulary. Because the study's goal was to examine individual differences, the same order of administration was used for all students to ensure that any potential order or fatigue effects would be equivalent throughout the sample.

Adjustment of Scores for Grade Differences

Because grade (fourth or fifth) was significantly related to performance on most nonstandardized tests, these scores were converted to *z* scores based on the mean and standard deviation for all students at the child's grade. Although these *z* scores were used in the statistical analyses, descriptive statistics will be reported for raw scores.

Results

The analyses of data proceeded in six steps. First, reading skill deficits were identified by examining scores on reading and spelling measures. Second, participants were assigned to reading ability groups based on the presence-absence and type(s) of reading deficits that were identified. Third, these reading-ability groups were compared with regard to literacy skills and other characteristics. Fourth, reading status was examined in relation to educational history, on which basis the children with early- and late-emerging reading disabilities of each type were differentiated. Fifth, early- and late-emerging RD groups were compared on all measures that were taken. Last, a retrospective analysis of prior achievement scores for all groups was conducted. Prior to analysis, distributions for all variables were examined for skewness and outliers; no such threats to validity were observed.

Identification of Reading Deficits

A child was considered to have a reading comprehension deficit if his or her standard score for Reading Comprehension was 86 or less. Because word-level processing deficits might show up in various guises in fourth and fifth grades, six measures were examined: (a) standard scores for Word Identification and Word Attack and (b) grade-adjusted *z* scores for the two spelling and two reading speed measures. Two composite scores were created: a speed factor, created by averaging the *z* scores for Word Reading Speed and Pseudoword Reading Speed, and an accuracy factor, derived by a principal components analysis of the other four scores. A child was considered to have a word-level deficit if one or more of the following criteria were met: a standard score of 86 or lower for Word Identification or Word Attack, an accuracy factor score below -0.90 , and/or a speed factor score below -0.90 .

Any cutoff point for separating deficient from nondeficient ranges within a continuously distributed skill dimension is necessarily an arbitrary one (Committee on the Prevention of Reading Difficulties in Young Children, 1998). In selecting our criteria for defining reading deficits, we were guided by research-based estimates that reading disabilities are usually seen in about 10%–20% of the school population in the United States (Myklebust & Boshes, 1969; Pennington, 1991; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). We also took into account the range of reading abilities that we observed in the ESI-P group, which we presumed to reflect the schools' judgment about what degree of severity was required for a child to qualify for supplemental services.

It also bears noting that deficits in text comprehension, word-level accuracy, and word-level speed were not identified by performance on a single reading task but rather on separate tests that were not scored against a common reference group. Hence, it cannot be assumed that the accuracy or speed of word-level processing on the comprehension test, or conversely the compre-

hension of items on word-level tests, would necessarily show the same deficiencies that were observed when these facets of reading were examined separately.

Assignment to Reading-Ability Groups

On the basis of their patterns of reading deficit(s), children were assigned to one of four groups. The RC group ($n = 12$) included those who met the criterion for a reading comprehension deficit but did not have a word-level deficit; in order to eliminate children with borderline word-level skills from this group (as recommended by Shankweiler et al., 1999), we also required scores above 90 on Word Identification and Word Attack. The WL group ($n = 28$) was made up of those children who met criteria for a word-level deficit but not for a reading comprehension deficit; to eliminate children with borderline comprehension skills, we also required a Reading Comprehension score above 90. Children in the WL-RC group ($n = 26$) met criteria for both word-level and comprehension deficits or were deficient in one area and borderline in the other. Last, if criteria for assignment to one of the three RD groups were not met, the child was considered not to have RD and was assigned to the NRD group ($n = 95$).

In all, there were 66 children with a reading disability of some type and 95 normally achieving readers in the sample. Within the RD and NRD groups, there were similar proportions of boys (47% and 50%, respectively), $\chi^2(1, N = 161) = 0.10, p = .76, \phi = .03$, and of fourth graders (49% and 58%, respectively), $\chi^2(1, N = 161) = 1.39, p = .24, \phi = .24$. Reading disability was also not reliably associated with socioeconomic differences among schools; 55% of RD and 45% of NRD students were from the lower SES schools, $\chi^2(1, N = 161) = 1.34, p = .25, \phi = .09$.

Comparing the RC, WL, and WL-RC types of RD, the proportions of boys (58%, 40%, and 50%, respectively) and the proportions of students from lower SES schools (42%, 54%, and 62%, respectively) did not differ significantly: for gender, $\chi^2(2, N = 66) = 1.38, p = .50, \phi = .15$; for SES, $\chi^2(2, N = 66) = 1.33, p = .52, \phi = .14$. Although all scores had been adjusted for grade differences prior to assigning students to reading ability groups, there were relatively more fourth graders with the WL type (71%) than with the RC (42%) and WL-RC (35%) types, $\chi^2(2, N = 66) = 7.88, p = .02, \phi = .35$.

Comparisons Among Reading-Ability Groups

Performance of the four reading-ability groups was compared for reading and spelling scores, reading-related cognitive and language measures, and questionnaire responses. Descriptive statistics for all of these variables are provided in Table 1. Group differences were examined using one-way multivariate analyses of variance (MANOVAs) with planned contrasts between the NRD group and each RD group. A separate MANOVA was conducted for each of the following sets of conceptually related dependent variables: the seven word-level literacy skills listed first in Table 1, Wilks's $\lambda = .37, F(21, 434) = 8.68, p < .01$; Listening Comprehension and Vocabulary, $\lambda = .83, F(6, 310) = 5.05, p < .01$; Phonological Awareness and the three Rapid Serial Naming measures, $\lambda = .68, F(12, 408) = 5.43, p < .01$; and the five self-efficacy ratings, $\lambda = .70, F(15, 423) = 3.97, p < .01$. Reading Comprehension, FSIQ, Print Exposure, Block Design, and Visual-

Table 1
 Performance on All Measures by Groups With No Reading Difficulties (NRD) and by Those With Deficits in Reading Comprehension (RC), Word-Level (WL) Processing, or Both (WL-RC)

Measure	NRD (n = 95)		RC only (n = 12)		WL only (n = 28)		WL-RC (n = 26)		η^2	Differences ^a
	M	SD	M	SD	M	SD	M	SD		
Literacy skills ^b										
Word Identification (SS)	111.20	11.60	107.90	13.80	95.10	5.60	90.80	5.80	.42***	1, 2 > 3, 4
Word Attack (SS)	105.70	12.20	106.90	16.00	84.80	6.90	83.20	9.40	.46***	1, 2 > 3, 4
Reading speed										
Words/second	1.87	0.31	1.78	0.19	1.41	0.29	1.39	0.28	.38***	1, 2 > 3, 4
Pseudowords/second	1.20	0.27	1.15	0.19	0.72	0.20	0.71	0.23	.47***	1, 2 > 3, 4
Exception word spelling	0.56	0.12	0.57	0.19	0.37	0.18	0.33	0.13	.34***	1, 2 > 3, 4
Pseudoword spelling										
Phonological score	0.80	0.19	0.70	0.24	0.46	0.27	0.42	0.28	.36***	1, 2 > 3, 4
Orthographic score	0.74	0.13	0.68	0.13	0.55	0.15	0.55	0.13	.27***	1, 2 > 3, 4
Reading Comprehension (SS)	107.20	12.70	79.30	4.30	103.10	8.30	78.00	9.30	.54***	1, 3 > 2, 4
Cognitive skills										
Listening Comprehension (SS)	112.70	12.90	105.90	8.70	110.20	14.50	99.90	15.80	.11***	1, 3 > 4; 1 > 2
WISC-III Vocabulary (SS)	11.60	2.40	10.00	2.70	11.00	2.50	8.70	2.90	.16***	1, 3 > 4
Phonological awareness	0.71	0.27	0.70	0.29	0.43	0.29	0.45	0.24	.18***	1, 2 > 3, 4
Rapid serial naming										
Letters/second	2.30	0.44	2.20	0.25	1.89	0.40	1.90	0.44	.17***	1, 2 > 3, 4
Digits/second	2.37	0.43	2.30	0.31	1.97	0.43	1.95	0.37	.17***	1, 2 > 3, 4
Objects/second	1.16	0.19	1.10	0.19	1.07	0.21	1.03	0.20	.06*	1 > 4
Visual-verbal learning	0.83	0.16	0.80	0.17	0.84	0.12	0.81	0.18	.01	
WISC-III Block Design (SS)	10.80	3.40	8.90	4.00	11.00	4.40	9.10	2.90	.06*	1 > 4
Full-Scale IQ estimate	107.10	13.40	97.50	17.00	105.80	16.00	93.50	12.30	.12***	1, 3 > 4
Questionnaire responses										
Print exposure (A')	0.83	0.06	0.82	0.07	0.81	0.06	0.78	0.10	.09**	1 > 3, 4; 2 > 4
Self-efficacy (0-7 scale)										
Reading ability	5.10	1.10	4.90	1.10	4.00	1.30	3.90	1.50	.16***	1, 2 > 3, 4
Reading speed	4.90	1.30	5.10	1.20	3.90	1.70	3.40	1.50	.16***	1, 2 > 3, 4
Spelling ability	5.00	1.40	4.90	1.70	3.90	1.20	3.90	1.50	.13***	1, 2 > 3, 4
Math ability	4.60	1.30	5.30	1.50	4.90	1.50	5.00	1.50	.02	
Intelligence	4.90	1.30	4.80	1.10	4.80	1.10	4.60	0.80	.01	

Note. Unless otherwise noted, means for raw scores (proportion correct) are listed. SS = standard (or scaled) score; WISC-III = Wechsler Intelligence Scale for Children—Third Edition.

^a 1 = NRD, 2 = RC only, 3 = WL only, 4 = WL-RC. ^b These eight measures were used to identify the reading deficits on which group assignments were based. By definition, a group with a given type of deficit will differ from a group that does not share that deficit on any measure(s) of the deficient skill. Two groups with the same type of deficit could differ in its severity, however, and only those comparisons are meaningful for this set of measures. * $p < .05$. ** $p < .01$. *** $p < .001$.

Verbal Learning measures were analyzed individually. Post hoc Tukey tests were conducted to clarify significant group effects. Table 1 shows the results of univariate tests for all measures, including omnibus effect sizes and significant pairwise differences between group means.

The two groups with word-level deficits (WL and WL-RC) performed more poorly than the other groups but did not differ reliably from each other on all tests that had been used to identify deficits in the accuracy or speed of word-level reading and spelling. They also differed on measures of two skills that have often been associated with impaired word-level processing: phonological awareness and rapid serial naming. This pattern was also seen for self-efficacy ratings of literacy skills, but there were no group differences in self-rated math ability or intelligence relative to peers. Effect sizes (d) for the observed skill deficits (relative to the

NRD group) in the WL and WL-RC groups, respectively, were as follows: 1.21 and 1.51 for Word Identification, 1.37 and 1.47 for Word Attack, 1.19 and 1.45 for word reading speed, 1.39 and 1.56 for pseudoword reading speed, 0.98 and 1.42 for exception word spelling, 1.21 and 1.37 for phonological spelling of pseudowords, 1.12 and 1.14 for orthographic spelling, 0.84 and 0.95 for phonological awareness, 0.74 and 0.88 for letter naming, 0.74 and 0.97 for digit naming, and 0.32 and 0.66 for object naming.

By definition, the two groups with comprehension difficulties (RC and WL-RC) had lower reading comprehension scores than other groups. These groups did not differ from each other in the severity of their reading comprehension deficits. Expected decrements in listening comprehension and vocabulary skills were seen, although the vocabulary deficit was not significant ($p = .07$) for

the RC group, whose small sample size afforded low statistical power. Results for Block Design and FSIQ were similar but only for the children with mixed deficits were mean scores significantly lower than in the NRD group. In addition, the WL-RC group's print exposure score was lower than that of the NRD and RC groups. Effect sizes (d) for the observed skill deficits (relative to the NRD group) in the RC and WL-RC groups, respectively, were as follows: 1.71 and 1.80 for Reading Comprehension, 0.48 and 0.91 for listening comprehension, 0.60 and 1.09 for Vocabulary, 0.53 and 0.59 for Block Design, 0.65 and 0.91 for estimated IQ.

Relation of Reading Deficits to Educational History

The top portion of Table 2 provides a breakdown of the sample according to their reading-deficit group assignments (based on research criteria) and their educational history classifications, defined earlier. The research-identified and school-identified groups overlapped considerably but far from perfectly. Most nontransient school-identified poor readers also met our criteria for RD, the exceptions being 9 (21%) of the 42 students in the ESI-P group and 4 (31%) of the 13 in the LSI group. Of the 66 participants who met the research criteria for WL, RC, or mixed deficits, 22 (33%) had not been considered by their schools to have reading problems.

Early-Identified Versus Late-Identified Reading Disabilities

The 66 students with RD by our criteria were next grouped according to time of identification, as summarized in the lower portion of Table 2. The 35 early-identified cases were those from the persistent and transient ESI subsamples. The 31 late-identified cases were from the LSI, PAR, and NOH groups. The composition of the early- and late-identified groups was similar with regard to the following: the proportion of boys (54% and 52%, respectively),

$\chi^2(1, N = 66) = 0.05, p = .83, \phi = .03$; the proportion of fourth graders (51% and 51%, respectively), $\chi^2(1, N = 66) = 0.01, p = .99, \phi = .01$; and the proportion from lower SES schools (57% and 52%, respectively), $\chi^2(1, N = 66) = 0.20, p = .81, \phi = .06$.

We had planned to conduct comparisons of the early-identified and late-identified cases within each reading-disability type. However, the RC group had insufficient numbers of cases (2 early and 10 late) to permit any meaningful statistical analyses. Therefore, only the groups with word-level deficits (WL and WL-RC) were examined for skill differences in relation to time of identification. For these comparisons, the two deficit groups were combined ($n = 54$) because so few differences between them had been found in the analyses described earlier.

Table 3 provides a comparison of the performance profiles of the 33 early-identified versus the 21 late-identified cases with word-level processing deficits. Means were very similar for most word-level literacy skills, cognitive abilities, print exposure, and self-evaluations. The late-identified cases, however, showed some indications that their deficits were less severe. They were faster than early-identified cases at reading both words and pseudowords in isolation ($ds = .75$ and $.70$, respectively) but still quite slow in comparison with unimpaired classmates. Their naming speeds were also somewhat slower than those of early-identified cases; although none was itself statistically significant, these differences were consistent across letter, digit, and object arrays ($ds = .35$ to $.55$). The late-identified group also spelled more exception words correctly ($d = 0.85$). Consistent with this, they were also better ($d = 0.46$), but not significantly so, at spelling pseudowords in orthographically appropriate ways.

Retrospective Analyses of Third-Grade School-Administered Achievement Tests

As noted earlier, the most recent standardized test scores available from school records were from third grade, 1–2 years prior to

Table 2
Distribution (Frequencies and Percentages) of Reading Status and Deficit Types in Relation to Educational History and Time of First Identification of Reading Disability (RD)

Group	N	Reading status		Reading deficit type		
		NRD	RD	WL only	RC only	WL-RC
Total sample	161	95 (59)	66 (41)	28 (17)	12 (7)	26 (16)
Educational history						
Early school-identified: Persistent	42	9 (21)	33 (79)	15 (36)	2 (5)	16 (38)
Early school-identified: Transient	10	8 (80)	2 (20)	2 (20)	0	0
Late school-identified RD	13	4 (31)	9 (69)	3 (23)	0	6 (46)
Parent concern only	13	9 (69)	4 (31)	1 (8)	1 (8)	2 (15)
No history of concern	83	65 (78)	18 (22)	7 (8)	9 (11)	2 (2)
Time of RD identification						
Early (Grades 1–3)			35	17	2	16
Late (Grades 4–5)			31	11	10	10

Note. Numbers in parentheses represent percentages. NRD = no reading difficulties; WL = word level; RC = reading comprehension.

Table 3
Comparison of Early-Emerging (First-Third Grades) and Late-Emerging (Fourth-Fifth Grades) Cases With Word-Level or Mixed Reading Deficits

Measure	Early emerging (<i>n</i> = 33)		Late emerging (<i>n</i> = 21)		<i>t</i>	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Literacy skills							
Word Identification (SS)	92.30	6.90	94.20	4.40	-1.25	.216	0.13
Word Attack (SS)	83.50	8.80	84.80	7.20	-0.57	.573	0.08
Reading speed (items/second)							
Words	1.29	0.22	1.55	0.28	-3.58	.001	0.75
Pseudowords	0.63	0.18	0.85	0.20	-4.18	< .001	0.70
Exception word spelling	0.29	0.15	0.44	0.13	-3.70	< .001	0.86
Pseudoword spelling							
Phonological score	0.42	0.29	0.47	0.25	-0.61	.543	0.16
Orthographic score	0.52	0.18	0.59	0.12	-1.86	.069	0.46
Reading Comprehension (SS)	91.30	16.40	90.60	14.10	0.17	.862	0.06
Cognitive skills							
Listening Comprehension (SS)	102.50	14.34	109.50	17.50	-1.53	.134	0.47
WISC-III Vocabulary (SS)	9.60	3.10	10.40	2.40	-1.07	.289	0.05
Phonological awareness	0.47	0.26	0.39	0.27	1.12	.267	0.30
Rapid serial naming							
Letters/second	1.83	0.36	2.00	0.49	-1.24	.222	0.35
Digits/second	1.88	0.35	2.10	0.44	-1.93	.061	0.50
Objects/second	1.01	0.18	1.12	0.23	-1.78	.083	0.55
Visual-verbal learning	0.81	0.15	0.85	0.15	-1.13	.263	0.29
WISC-III Block Design (SS)	10.80	3.90	9.10	3.60	1.65	.107	0.57
Full-Scale IQ Estimate	101.10	16.50	98.30	14.10	0.65	.518	0.18
Questionnaire responses							
Print exposure (<i>A'</i>)	0.80	0.07	0.79	0.10	0.32	.752	0.01
Self-efficacy (0-7 Likert scale)							
Reading ability	3.90	1.50	4.10	1.30	-0.53	.601	0.15
Reading speed	3.60	1.60	3.80	1.70	-0.55	.587	0.17
Spelling ability	3.80	1.40	4.00	1.20	-0.37	.717	0.09
Math ability	5.00	1.40	4.90	1.60	0.29	.774	0.09
Intelligence	4.60	1.00	4.90	1.00	-1.07	.290	0.28

Note. Unless otherwise noted, means for raw scores (proportion correct) are listed. WISC-III = Wechsler Intelligence Scale for Children—Third Edition; SS = standard (or scaled) score.

data collection for this study. National percentile scores for reading vocabulary (word recognition), reading comprehension, spelling, and math were analyzed to compare prior and current achievement of the various reading-ability groups and to examine the overall level of reading achievement in the sample versus the districts from which it was drawn.

Comparison of reading-ability groups on third-grade achievement tests. A one-way MANOVA, with the three reading and spelling scores as the dependent variables, was conducted to compare prior literacy achievement across six groups: NRD, RC, and early- and late-identified WL and WL-RC. As illustrated in Figure 1, there had been substantial group differences in third grade, Wilks's $\lambda = .50$, $F(15, 409) = 7.87$, $p < .01$, $\eta^2 = .21$. Univariate analyses indicated that there were reliable effects ($p < .01$) for all three measures: Reading Vocabulary ($\eta^2 = .36$), Reading Comprehension ($\eta^2 = .45$), and Spelling ($\eta^2 = .22$). Post hoc Tukey tests revealed no significant differences in third grade among the NRD, RC, and late-identified

WL groups on any measure (d s = .02 to .63), nor among the other three groups (d s = 0.01 to 0.75), except for a lower reading comprehension mean for early- than late-identified cases with mixed deficits ($d = 0.97$). For all three measures, the NRD group's mean exceeded those of the three lowest achieving groups (d s = 0.88 to 1.92); the RC group did better than the two early-identified groups (d s = 0.83 to 1.58); and the WL-late group outperformed the WL-RC-early group (d s = 0.97 to 1.38).

There were also a few differences among the six groups in third-grade math achievement, which was examined in a separate ANOVA, $F(5, 150) = 7.83$, $p < .01$, $\eta^2 = .21$. The mean percentile for the WL-RC-early group (27.9, $SD = 17.9$) was lower than those of the RC (58.5; $SD = 31.7$), NRD (63.3; $SD = 24.4$), and WL-late (68.7; $SD = 19.2$) groups ($d = 1.16$ to 1.55) and was also somewhat lower ($d = 0.59$ to 0.70), but not significantly, than the means for the WL-early (46.4; $SD = 24.6$) and the WL-RC-late (43.6; $SD = 18.6$) groups. The other five

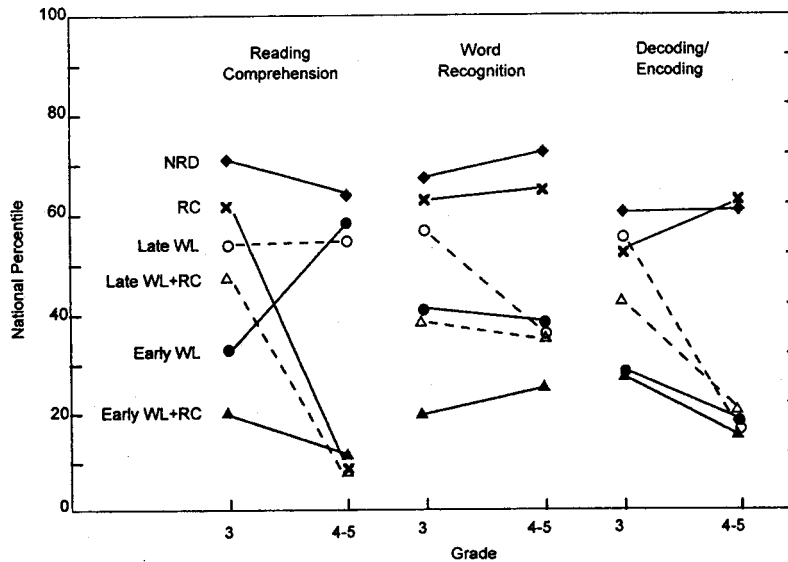


Figure 1. Third-grade versus fourth- and fifth-grade achievement on reading comprehension, word recognition, and decoding-spelling measures for the early- and late-identified reading-ability groups. Scores on school-administered tests (Comprehensive Test of Basic Skills or Iowa Test of Basic Skills) in third grade are contrasted with their closest counterparts among the individually administered research measures (Peabody Individual Achievement Test—Revised or Woodcock—Johnson Psychoeducational Battery—Revised) in fourth or fifth grade. NRD = no reading difficulties; RC = reading-comprehension deficit; WL = word-level deficit.

groups did not differ reliably from each other in third-grade math achievement.

To examine the past versus current literacy skills of the groups, we looked at the third-grade scores in comparison with national percentile scores on the individual tests in our battery that tapped reasonably similar skills: PIAT Reading Comprehension, WJ-R Word Identification, and (because we used no standardized spelling measure) WJ-R Word Attack. Although comparisons across tests that have not been conformed can be questionable, the pattern of findings (shown in Figure 1) suggests that in this instance the two sets of scores are sufficiently comparable for our purposes. A mixed model ANOVA was conducted with grade (third vs. fourth-fifth) and reading skill (comprehension, word recognition, decoding-encoding) as the within-subject repeated measures and group as the between-subjects factor. Because the question of interest was whether there had been a loss (or gain) over time in the achievement levels of the RD groups on any aspect of reading, only the three-way interaction effect was tested, $F(10, 300) = 10.00, p < .01, \eta^2 = .25$.

Post hoc correlated t tests indicated that the following changes between third-grade and fourth- and fifth-grade achievement were strong and reliable: the declines in reading comprehension percentiles in the RC group (53 points; $p < .01$) and the late-identified WL-RC group (36 points; $p < .01$); the 20-point decrease in word recognition in the WL-late group ($p < .01$); the drops in decoding-encoding scores in the WL-late group (40 points; $p < .01$), the WL-RC-late group (22 points; $p = .04$), and the WL-RC-early group (14 points; $p = .01$); and the 24-point rise for reading comprehension in the WL-early group ($p = .01$). These shifts resulted in increased dissociations between word-level and comprehension components of reading over time. Reflecting this

change, the correlation between word recognition and reading comprehension scores was much higher in third grade ($r = .75$) than subsequently (.48) in the sample ($z = 4.21, p < .01$).

Comparison of reading achievement in the sample with district records. Because this was a sample ascertained through voluntary participation and one in which children with reading difficulties were intentionally overrepresented, it would not be appropriate to view it as population representative. It was of interest to know, however, how the sample's overall reading level compared with that of the districts from which it was drawn. Although contemporaneous data were not available, district performance on the third-grade tests could be examined for this purpose. In the first district, which administered the ITBS, the mean national percentile for Total Reading was 55.13 ($SD = 28.12$) in our sample. In comparison, district averages for the 2 years during which children in the sample had taken the test were 53 and 54, according to information provided by the superintendent's office. In the second district, which used the CTBS, local percentile scores were on file for children in the sample. Their mean score of 49.63 ($SD = 29.80$) was almost identical to the district average of 50.00. Hence, for children from both districts, the average reading level in the sample was similar to what one would expect if the sample had been randomly drawn from this population.

Discussion

The findings provide some answers to each of the questions posed initially about children with late-identified reading disabilities: How heterogeneous are they with regard to the type of deficit(s)? Compared with their early-identified peers, do late-identified disabled readers have less severe, less broad, or alto-

gether different profiles of reading achievement and other characteristics? Does late identification of the reading problem occur because schools have overlooked weaknesses that were evident earlier, or because their reading difficulties actually emerge later?

Heterogeneity of Late-Identified Reading Disabilities

One hypothesis that we entertained was that most fourth and fifth graders with late-identified RD would be characterized primarily by comprehension-based weaknesses, as is commonly believed. Alternatively, we wondered whether many of these students might instead (or also) have lower level processing problems similar to those typically seen among younger children with reading disabilities, an observation that has been anecdotally reported to us by teachers and clinicians. In our sample, we found that the late-identified cases were not a homogeneous group. Rather, 35% had word-level processing deficits in combination with adequate comprehension skills, 32% showed weak comprehension skills accompanied by good lower level skills, and 32% exhibited both kinds of difficulty. In comparison, the distribution of these three types was more uneven among the children whose reading problems were identified early: 49%, 6%, and 46%, respectively. Recall, however, that deficits in text comprehension, word-level accuracy, and word-level speed were identified on the basis of performance on separate tests with scores based on different reference samples, so these percentages might differ somewhat if all components were measured using a common task.

The fact that two thirds of the late-identified cases met our criterion for a comprehension deficit is consistent with the widely held view that many struggling readers at this age have difficulty in understanding written material. On the other hand, we also found word-level processing deficits in 64% of these students, supporting the notion that word recognition, decoding, and spelling also continue to pose challenges to progress in reading achievement beyond the primary grades.

Although a third of the group exhibited both kinds of weaknesses, those who met our criteria for a deficit in just one major component of reading were relatively unimpaired in the other. We had anticipated that because top-down and bottom-up skills are thought to interact facilitatively during reading (Committee on the Prevention of Reading Difficulties in Young Children, 1998), poor comprehension would drag down word recognition and that poor word recognition would impair text comprehension, yielding relatively flat skill profiles of the sort seen in third grade. In fourth and fifth grade, however, there was only a weak tendency, if any, of this sort. Instead, proficiency in the unimpaired component was about the same as in the normally achieving group. Furthermore, even the early-emerging group with word-level difficulties exhibited surprising strength in comprehension with the higher grades, suggesting that their low reading comprehension scores in third grade were attributable primarily to word recognition limitations that were an impediment to successful text comprehension at that point in their development.

Cognitive Skills and Other Characteristics Associated With Reading Deficit Types

In this sample of fourth and fifth graders, reading difficulties were accompanied by several other characteristics that have often

been observed in prior studies, primarily with younger students but also with adolescent and adult samples. As one would expect, these associated weaknesses differed according to whether reading comprehension or word recognition was impaired. We will summarize these profiles of the reading-deficit groups here and then discuss differences between early- and late-emerging reading disabilities in the next section.

Comprehension-only type. The children with poor reading comprehension but strong word recognition, decoding, and spelling skills resemble the picture of a "pure" comprehension deficit that has been described in previous research (Guthrie, 1973; Nation, 1999; Stothard & Hulme, 1992; Yuill & Oakhill, 1988, 1991). This group's mean Vocabulary, Listening Comprehension, Block Design and FSIQ scores were also lower than those of normally achieving students, although these differences were not as large as the gap in reading comprehension itself. Although not statistically significant (perhaps reflecting the low power afforded by the small size of the RC group), the consistency and strength of these findings converge with results of prior research in suggesting that successful reading comprehension depends on a host of factors, including oral language proficiency, breadth of vocabulary, organizational skills, and general reasoning abilities. As reviewed in the recent Rand report (Snow, 2002), precisely how weaknesses in these areas contribute to failures of comprehension is not yet well understood.

Word-level-only type. Other students fit the typical "dyslexic" profile outlined in the literature (Bruck, 1990; Nation, 1999; Shaywitz, 1996). That is, they had comprehension skills comparable with those of nondisabled readers but exhibited deep and broad bottom-up processing deficits in word and pseudoword reading (accuracy and speed) and spelling and in phonological awareness and naming speed. Print exposure was also lower than for normally achieving students, and these children rated themselves lower in literacy skills than did the better readers, as has often been reported (Chapman, 1988; Schunk, 1989; Simpson, Licht, Wagner, & Stader, 1996). This group was unimpaired in other respects, including listening comprehension, vocabulary, and IQ. Bruck (1990), who observed some of these phenomena in older dyslexics, has suggested that individuals with this type of RD are able to succeed in reading comprehension by using their otherwise strong cognitive and linguistic abilities to compensate for word-level processing difficulties, probably by using contextual cues to make educated guesses that, often enough, are sufficiently accurate for an understanding of the material to be gained.

Mixed type. The remainder of the children with RD had difficulties with both word-level and higher order literacy skills, and these deficits were about equal in severity to those of the children whose difficulties were confined to only one major component of reading. Consistent with this, they exhibited the same sorts of additional weaknesses that were observed in cases with only bottom-up difficulties (phonological awareness, rapid naming, print exposure) and in those with only top-down impairments (listening comprehension, vocabulary, and nonverbal reasoning). Because their word-level reading skills were about as poor as those of the children in the WL group, who comprehended text successfully, the reading comprehension difficulties of the group with mixed deficits probably did not stem solely from inefficient word recognition but also from vocabulary limitations, insufficient

knowledge of language structure, lack of background knowledge, and weak inferential abilities.

Characteristics Associated With Early- Versus Late-Identified Reading Disabilities

In our sample and those of other researchers (Shankweiler et al., 1999; Yuill & Oakhill, 1991), few children with early-identified RD showed the comprehension-only deficit pattern, suggesting that this type occurs most often as a late-emerging reading problem. It is generally held that the roots of reading comprehension difficulties are weaknesses (lexical, syntactic, conceptual, inferential, organizational, etc.) originating in the preschool and early school years. In the primary grades, such limitations (unless especially severe) appear not to hinder the learning of word recognition skills nor performance on conventional "reading comprehension" measures, which rarely demand high levels of vocabulary, reasoning, or background knowledge. Indeed, the RC group in our sample was nearly as high achieving as the NRD group on the third-grade tests administered by their schools. Hence, it is not terribly surprising that only 2 of these children had yet come to the attention of their teachers, despite the marked reading comprehension weaknesses that they subsequently exhibited on our tests. It is interesting, furthermore, that their high ratings of their own abilities indicated that the students themselves were unaware that their understanding of both oral and written material was weaker than that of their classmates.

In contrast, a majority (61%) of the children with each of the other two types of reading disability—word level only and mixed—were identified by third grade or sooner. We had wondered whether late-identified cases might have less severe (but otherwise similar) deficits than early-identified poor readers, or whether they might even show qualitatively different patterns of achievement and characteristics than the early-identified poor readers. Somewhat to our surprise, the data indicated that two groups differing in time of identification were quite similar in their profiles, providing only limited support for the severity hypothesis. That is, the early- and late-identified groups had about equally large deficits in many word-level reading abilities (including accuracy of word recognition and pseudoword decoding and spelling) as well as in the related skills of phonological awareness and naming speed. Likewise, on the questionnaires, they were similar in print exposure (title recognition) and self-efficacy ratings. Of the skills assessed, only reading speed (for both words and pseudowords) and spelling of exception words differed significantly, with the late-identified cases performing better than the early-identified ones, although still substantially less well than normally achieving classmates.

Explanation for Late Identification

Identification of RD occurred after third grade or had not yet been made by the schools for 31 students in this sample who met research criteria for reading deficits in fourth or fifth grade. Was it the case that these students had earlier deficiencies that were overlooked in the primary grades, perhaps because these children displayed high intelligence, good behavior, or compensatory strategies? Or did these children actually not begin to have detectible difficulties until after third grade? The comparison of third grade

with fourth- and fifth-grade achievement levels in Figure 1, although not definitive, is more consistent with the latter explanation. Achievement by the best (NRD) and poorest (early-identified WL-RC) readers was quite stable over tests and grades. Scores of the other groups, although consistent across reading components in the third grade, showed greater disparities between aspects of reading proficiency in fourth or fifth grade. The RC group maintained its high degree of skill in word recognition and decoding-encoding but fell sharply in reading comprehension. Comprehension levels also dropped for the late-identified WL-RC group and were accompanied by smaller decreases in other aspects of reading proficiency. The late-WL group remained strong in comprehension but declined sharply in word-level processing skills, bringing them down to the level of the early-WL group, which actually gained in comprehension while remaining weak in word recognition and decoding. The available data are thus not consistent with the notion that the schools had overlooked or disregarded earlier signs of reading difficulties in the late-identified cases. In short, these children's reading disabilities appear not just to be late identified but actually late emerging.

Prevalence of Late-Emerging Reading Disabilities

How common are late-emerging disabilities? Prevalence statistics for RD necessarily vary according to the criteria used to define it. Therefore, a more sensible question is, What proportion of all children with RD (defined by a given set of criteria) are late identified? Of the 66 fourth and fifth graders who met our criteria for some type of RD, 31 (47%) were late identified (see Table 2). However, it must be recalled that we intentionally sampled only 50% of the available participants with no history of reading problems and of those with a history of early school-identified RD. If all such cases had been tested, by extrapolation one might expect the proportion of late-emerging cases to drop to about 41% of all students with RD.

This was not an epidemiological sample but a self-selected one, however, so the estimated prevalence of late-identified RD based on this study is not necessarily representative of the proportion of such students in the general population. Some prevalence estimates for late-emerging RD are available, however, from epidemiological studies. Kavale and Reese (1992) examined a stratified sample of 1,070 Iowa students with learning disabilities. They found that of all initial referrals through fifth grade, only 22% were made after third grade. Note, however, that this study did not separate RD from other kinds of learning disabilities and only included school-identified cases, which have previously been shown to exclude many children whose deficient reading skills are revealed on tests administered by researchers (e.g., Scarborough, 1989; Shaywitz et al., 1990). In two other studies, RD determinations were based on research assessments, and higher prevalence estimates were obtained.

In the Connecticut Longitudinal Study (Shaywitz et al., 1992), the reading status of 414 students in a population-representative sample was compared at third versus fifth grade, using an achievement-aptitude discrepancy as the basis for identifying RD cases. Of all fifth graders with RD, 42% were late-emerging cases (i.e., who met the classification criterion in fifth grade but were normally achieving readers in third grade). Similarly, Badian (1999) obtained annual reading comprehension scores for all 1,008

children who were continuously enrolled in a single Massachusetts school district from first through eighth grade. Children were considered to have late-emerging RD if their mean standard score was below 90 over fifth through eighth grade but above 90 in earlier grades. This group constituted 46% of all of the children in fifth grade or higher who had reading comprehension difficulties.

Even though the Connecticut study used a discrepancy definition of RD (which we did not), and even though the Massachusetts study used only reading comprehension scores as the basis for defining RD, the estimated proportion of late-emerging cases among older elementary and middle school students in these samples (42% and 46%) is quite similar to ours (41%). Taken all together, the available data suggest that among children of this age with reading disabilities, a considerable proportion are late-emerging cases whose weaknesses were not identifiable from performance on conventional reading tests during the primary grades.

Educational Implications

The converging findings from this study and others have implications for the way that reading difficulties are assessed and treated in schools, particularly after the primary grades. First, it appears that the late-emerging reading disabilities of many children go undetected by the schools, at least for the first year or two after their emergence. To ensure that these students receive assistance as soon as possible, school personnel should heighten their alertness to the possibility that successful early achievers may undergo abrupt declines in performance relative to classmates.

Second, reading disabilities in children beyond the primary grades appear to be heterogeneous with regard to the nature of their skill deficits, as has been previously shown for adolescent and adult samples. Some children have comprehension problems only, some have just word-level difficulties, and some exhibit across-the-board weaknesses. Hence, both assessment and instruction need to be aligned with this reality. Among children who earn low scores on reading comprehension tests, for instance, many are likely to have word-level processing weaknesses. Distinguishing them from students with comprehension difficulties alone, and identifying cases with only lower-level reading deficits, should be a goal of assessment at this age. Most important, intervention programs then need to be selected on the basis of children's deficit type(s) rather than the overall grade level. To strengthen word-level processing, intensified and level-appropriate phonics-based instruction may be called for; to improve reading comprehension, many programs have been shown to be effective, and using a combination of these has been recommended by the National Reading Panel (2000).

Third, an increased emphasis is now being placed on the early identification and prevention of reading problems, and rightly so. Much is known about predicting reading disabilities in the primary grades from assessments made in kindergarten or earlier (Scarborough, 1998), and effective interventions to facilitate reading acquisition by children at risk for such difficulties have been developed (Committee on the Prevention of Reading Difficulties in Young Children, 1998). The findings of this study suggest that such prevention efforts, although laudable, will probably not serve to inoculate the student population against the occurrence of late-emerging reading disabilities.

Limitations of the Study and Directions for Future Research

The conclusions of the study cannot be considered definitive in view of the many limitations of this initial investigation of late-emerging RD. The participants were volunteers, and the extent to which this may have biased the findings cannot be estimated. Several subgroups were quite small in number, so that statistical comparisons of those groups with others were underpowered; hence, only tentative interpretations can be made of the several group differences that had large effect sizes but did not meet the significance criterion. Although our results converge well with those from prior research, particularly studies of adolescent and adult readers, replication of the present findings will be needed.

In addition, a broad range of measures was given to the sample on just one occasion, and the only longitudinal component of the study was a retrospective analysis of school-administered group achievement scores. Without more thorough assessments at younger ages, it was not possible to investigate several important questions about the development of children with late-emerging RD. For example, could the fourth-grade slump in achievement be predicted from performance measures in the primary grades? That is, decreases in word-level reading skills might be foreshadowed by slower-than-average processing speed, and declines in reading comprehension might be predictable from assessments of oral-language abilities, reasoning abilities, and comprehension strategy usage at younger ages. Another unanswered question is: To what extent might late-emerging reading difficulties result from instructional differences? Conceivably, excellent primary grade teaching might forestall the onset or identification of a reading problem, whereas insufficient early instruction might be largely responsible for deficient knowledge and skills seen in later grades, particularly with regard to comprehension abilities. To address these issues and to confirm and extend the findings of this preliminary study, a large-scale, prospective longitudinal investigation in a population-representative sample will be required.

References

- Anderson, R. C. (1980). Study strategies and adjunct aid. In R. J. Spiro, B. C. Bruce, & W. R. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, artificial intelligence linguistics, and education* (pp. 415-431). Hillsdale, NJ: Erlbaum.
- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Champaign, IL: Center for the Study of Reading and the National Academy of Education.
- Applebee, A. N., Langer, J. A., & Mullis, I. (1987). *Learning to be literate in America: Reading, writing, and reasoning*. Princeton, NJ: National Assessment of Educational Progress, Educational Testing Service.
- Badian, N. A. (1999). Reading disability defined as a discrepancy between listening and reading comprehension: A longitudinal study of stability, gender differences, and prevalence. *Journal of Learning Disabilities, 32*, 138-148.
- Bruck, M. (1990). Word recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26*, 439-454.
- Calfee, R., Lindamood, C., & Lindamood, P. (1973). *Lindamood Auditory Conceptualization Test*. Allen, TX: DLM Teaching Resources.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.

- Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.
- Chapman, J. W. (1988). Learning disabled children's self-concepts. *Review of Educational Research*, 58, 347-371.
- Chi, M. T. H., & Koeske, R. D. (1983). Network presentation of a child's dinosaur knowledge. *Developmental Psychology*, 19, 29-39.
- Committee on the Prevention of Reading Difficulties in Young Children. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Research Council.
- Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology*, 82, 733-740.
- Duke, N. K. (2000). For the rich it's richer: Print environments and experiences offered to first-grade students in very low- and very high-SES school districts. *American Educational Research Journal*, 37, 456-457.
- Dunn, L. M., & Marquardt, F. C. (1989). *Peabody Individual Achievement Test—Revised*. Circle Pines, MN: American Guidance Service.
- Fowler, A. E., & Scarborough, H. S. (1993). *Should reading disabled adults be distinguished from other adults seeking literacy instruction? A review of theory and research*. (Technical Report No. 93-6). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.
- Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology*, 65, 294-299.
- Hare, V. C., & Pulliam, C. A. (1980). College student's metacognitive awareness of reading behavior. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives on reading research instruction. Twenty-ninth yearbook of the National Reading Conference* (pp. 226-271). Washington, DC: National Reading Conference.
- Hayes, D. P., & Ahrens, M. (1988). Vocabulary simplification for children. *Journal of Child Language*, 15, 395-410.
- Juel, C. (1991). Beginning reading. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 749-788). New York: Longman.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kavale, K. A., & Reese, J. H. (1992). The character of learning disabilities: An Iowa profile. *Learning Disability Quarterly*, 15, 74-94.
- Kletzein, S. B. (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly*, 26, 67-85.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Lindamood, C., & Lindamood, P. (1979). *Lindamood Auditory Conceptualization Test*. Austin, TX: Pro-Ed.
- McDowell, J. A., Schumm, J. S., & Vaughn, S. (1992, November). *Assessing exposure to print: Development of a measure for primary children*. Paper presented at the meeting of the National Reading Conference, San Antonio, TX.
- Myklebust, H. R., & Boshes, B. (1969). *Minimal brain damage in children* (Final Rep. No. 108-65-142). Washington, DC: U.S. Department of Health, Education, and Welfare, Neurological and Sensory Disease Control Program.
- Nation, K. (1999). Reading skills in hyperlexia: A developmental perspective. *Psychological Bulletin*, 125, 338-355.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67, 359-370.
- Nation, K., & Snowling, M. (1998). Semantic processing and the development of word recognition: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39, 85-101.
- National Reading Panel. (2000). *Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: National Institutes of Health.
- Pearson, P., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, 11, 201-209.
- Pennington, P. F. (1991). *Diagnosing learning disorders: A neuropsychological framework*. New York: Guilford.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Sattler, J. (1992). *Assessment of children: WISC-III and WPPSI-R supplement*. San Diego: Author.
- Scarborough, H. S. (1984). Continuity between childhood dyslexia and adult reading. *British Journal of Psychology*, 75, 329-348.
- Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology*, 81, 101-108.
- Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-119). Timonium, MD: York Press.
- Schunk, D. H. (1989). Self-efficacy and cognitive achievement: Implications for students with learning problems. *Journal of Learning Disabilities*, 22, 14-22.
- Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., et al. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, 3, 69-94.
- Shaywitz, S. E. (1996). Dyslexia. *Scientific American*, 275, 98-104.
- Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, 326, 145-150.
- Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association*, 264, 998-1002.
- Simpson, S. M., Licht, B. G., Wagner, R. K., & Stader, S. R. (1996). Organization of children's academic ability-related self-perceptions. *Journal of Educational Psychology*, 88, 387-396.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34-50.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Arlington, VA: Rand Corporation.
- Spear-Swerling, L. (2001, June). *Fourth-graders' performance on two different measures of reading comprehension*. Paper presented at the meeting of the Society for the Scientific Study of Reading, Boulder, CO.
- Spring, C., & French, L. (1990). Identifying children with specific reading disabilities from listening and reading discrepancy scores. *Journal of Learning Disabilities*, 23, 53-58.
- Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing: An Interdisciplinary Journal*, 4, 245-256.
- Strucker, J. (1995). *Patterns of reading in adult basic education*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Swanson, H. L., Cochran, K. F., & Ewers, C. A. (1990). Can learning disabilities be determined from working memory performance? *Journal of Learning Disabilities*, 23, 59-67.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Harcourt Brace Jovanovich.
- White, K. B. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91, 461-481.

- Wilson, A. M., & Lesaux, N. K. (2001). Persistence of phonological processing deficits in college students with dyslexia who have age-appropriate reading skills. *Journal of Learning Disabilities, 34*, 394-400.
- Wong, B. Y. L., & Wong, R. (1986). Study behavior as a function of metacognitive knowledge about critical task variables: An investigation of above average, average, and learning disabled readers. *Learning Disabilities Research, 1*, 101-111.
- Woodcock, R. W., & Johnson, M. B. (1989). *The Woodcock-Johnson Psychoeducational Battery—Revised*. Itasca, IL: Riverside.
- Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension [Special issue]. *Applied Cognitive Psychology, 2*, 33-45.
- Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension*. Cambridge, England: Cambridge University Press.

Received July 22, 2002

Revision received November 5, 2002

Accepted November 8, 2002 ■