

Early Reading Development in Children at Family Risk for Dyslexia

Bruce F. Pennington and Dianne L. Lefly

In a 3-year longitudinal study, middle- to upper-middle-class preschool children at high family risk (HR group, $N = 67$) and low family risk (LR group, $N = 57$) for dyslexia (or reading disability, RD), were evaluated yearly from before kindergarten to the end of second grade. Both phonological processing and literacy skills were tested at each of four time points. Consistent with the well-known familiarity of RD, 34% of the HR group compared with 6% of the LR group became RD. Participants who became RD showed deficits in both implicit and explicit phonological processing skills at all four time points, clearly indicating a broader phonological deficit than is often found at older ages. The predictors of literacy skill did not vary by risk group. Both risk groups underwent a similar developmental shift from letter-name knowledge to phoneme awareness as the main predictor of later literacy skill. This shift, however, occurred 2 years later in the HR group. Familial risk was continuous rather than discrete because HR children who did not become RD performed worse than LR non-RD children on some phonological and literacy measures. Finally, later RD could be predicted with moderate accuracy at age 5 years, with the strongest predictor being letter-name knowledge.

INTRODUCTION

Although there is extensive knowledge both about the phonological deficits found in cross-sectional studies of individuals with reading disability, or RD (Bradley & Bryant, 1978; Bruck, 1993; Liberman, 1973; Liberman, Rubin, Duques, & Carlisle, 1985; Pennington, Van Orden, Smith, Green, & Haith, 1990; Pratt & Brady, 1988) and about the longitudinal predictors of normal variations in reading skill (Bradley & Bryant, 1983; Bryant & Bradley, 1985; Jorm, Shore, MacLean, & Matthews, 1984; Mann & Liberman, 1984; Scarborough, 1998; Wagner & Torgesen, 1987; Wolf, 1991), there is considerably less information about the longitudinal predictors of reading skill in young children who will later become RD. Because of the robust familiarity and heritability of RD (DeFries, Fulker, & LaBuda, 1987; Hallgren, 1950; Pennington et al. 1991), longitudinal studies of young children at family risk for RD are quite feasible: Roughly between 30% and 50% of such children will become RD (Gilger, Pennington, & DeFries, 1991). Such studies can potentially address how genetic and environmental risk factors interact to produce RD, and whether the predictors of RD are similar to the predictors of normal variations in reading skill. The answers to these questions have important public policy implications because of the high and increasing demand for literacy in the global economy. Hence, early identification and treatment of children at risk for literacy problems is an important goal (Snow, Burns, & Griffin, 1998).

More generally, longitudinal studies of children at family risk for a disorder address a number of fundamental questions in developmental psychopathology (Plomin & Rutter, 1998), several of which can be

framed in terms of continuity vs discontinuity. The current study addresses three such questions about RD: (1) Is there continuity in the underlying phonological phenotype across development? (2) Do both extreme and normal variations lie on the same continuum such that they will have similar predictors? and (3) Is familial risk continuous or discrete? In addition to these three questions, the current study addressed a fourth question crucial for early identification: (4) How accurately can we identify at age 5 children who will later become RD? In the discussion that follows, these four questions are placed in the context of existing knowledge.

Continuity in the Underlying Phenotype in RD?

With regard to the first question, the cross-sectional studies cited earlier as well as other data suggest considerable continuity in the nature of the underlying phonological phenotype in RD. Using a reading level match design, these and other studies (e.g. Pennington, Cardoso-Martins, Green, & Lefly, in press) have consistently found that RD children, adolescents, and adults perform worse than younger normal readers matched on word recognition (who are termed reading-age, or RA, controls) on measures of both the phonological coding of written language, that is, nonword reading (see Rack, Snowling, & Olson, 1992 for a review) and phoneme awareness of spoken language. Thus, although groups with RD make developmental progress in these two skills, they do not reach RA

levels of proficiency. This developmental continuity in the underlying written and spoken language deficit in RD provides a straightforward explanation of the disorder, which is that individuals with RD are poor at word recognition because they have a deficient ability to decode unfamiliar words (the phonological coding deficit), and they are poor at phonological coding because they are poor at phoneme awareness, which is necessary for mastering an alphabetic orthography. Further support for this explanation is provided by the results of cross-sectional twin studies which have found that deficits in both phonological coding and phoneme awareness are heritable and share significant genetic covariance with each other and with reading deficits (Olson, Forsberg, & Wise, 1994; Olson, Wise, Connors, & Rack, 1989). The genetic risk factors for RD, therefore, appear to act, at least in part, by altering the development of phoneme awareness and phonological coding.

Besides this evidence for continuity, there is also evidence that the problem with phoneme awareness is fairly specific across ages. Cross-sectional studies of RD groups rarely have found differences relative to RA controls on measures of other, implicit, phonological processing skills such as speech perception, rapid serial naming, or verbal short-term memory (STM), although the RD groups in such studies differed from chronological age (CA) controls (e.g., Pennington et al., in press; Pennington et al., 1990). Because differences relative to CA controls on these measures of implicit phonological processing could be due to differences in reading experience or to other uncontrolled factors, the empirical support for a causal relation between a difficulty with phoneme awareness and reading difficulty is much stronger than it is for other phonological problems.

The impression of both continuity and specificity in the phonological phenotype in RD, however, could be misleading for several reasons. First, the CA differences on implicit phonological tasks could also suggest a broader phonological deficit. Second, the strength of the relation between phoneme awareness and reading deficits may partly derive from the fact that there is a reciprocal relation between reading and phoneme awareness (Morais, Cary, Algeria, & Bertelson, 1979), which could produce a positive feedback loop such that practice in decoding unfamiliar words would enhance phoneme awareness, which in turn would enhance phonological coding ability. If individuals with RD relied less on phonological coding as a reading strategy, they would benefit less from this feedback loop, even after controlling for word recognition level. Finally, the enormous developmental changes that occur in phonological and language de-

velopment in the preschool period make it very unlikely that continuity would extend downward to younger ages—even typically developing children do not develop phoneme awareness until around age 5 years (Yopp, 1988).

In sum, there is impressive evidence of continuity in the underlying phonological phenotype in RD in school-age and adult samples, but developmental and other considerations argue that at earlier ages a broader phenotype will be found. It is quite unlikely that the genetic variations that influence RD act directly on literacy or phoneme awareness because reading did not exist for most of human evolution. Instead, these genetic variations likely contribute to individual differences in basic language and cognitive skills, which in turn contribute to individual differences in phoneme awareness and literacy. Therefore, one of the questions of interest in this longitudinal study was whether the developmental continuity and specificity in the phonological phenotype of RD would extend to younger ages. As discussed in more detail later, evidence from other longitudinal studies of children at high family risk for RD suggests that it does not.

Do Normal and Extreme Variations in Reading Skill Have Similar Predictors?

Phoneme awareness predicts normal variations in later reading skill (e.g., Bradley & Bryant, 1983; Wagner & Torgesen, 1987), but so do other phonological and language skills (see Scarborough, 1998, for a review). The strongest predictor is letter-name knowledge, but it is unclear why this is the case. Several studies have shown that letter-name knowledge, besides being the most powerful predictor of later reading skill, also has a relationship to phonological awareness (Ehri & Wilce, 1985; Elbro, Borstrom, & Petersen, 1998; Gallagher, Frith, & Snowling, 2000; Johnston, Anderson, & Holligan, 1996; Read, Zhang, Nie, & Ding, 1986; Share, 1995). Letter-name knowledge is at the intersection between spoken and written language because letters are the written representations of phonemes or combinations of phonemes. It is plausible that the ability to learn letter names depends on underlying phonological development. As Share (1995) pointed out, letter names are, after all, nonwords, and the ability to repeat and remember nonwords has been shown to discriminate both children with specific language impairment (Bird, Bishop, & Freeman, 1995) and those with dyslexia (Brady, 1997) from controls.

Only a few studies have examined the predictors of extreme variations in reading skill in children at familial risk for dyslexia. Scarborough (1989, 1990,

1991), Gallagher et al. (2000), and Elbro et al. (1998) each conducted prospective studies in different countries in which both family risk and phonological processing were examined as predictors of reading outcome. All three studies found (1) a substantial familial risk for dyslexia, which affected 37% to 68% of the high-risk (HR) samples but only 2% to 12% of the low-risk (LR) samples; (2) a broader linguistic and phonological phenotype during the preschool period in the children who later became dyslexic than is typically found in school-age and adult samples; and (3) predictors of later literacy that were similar to those identified in longitudinal studies of normal reading development (i.e., letter-name knowledge, phoneme awareness, and other phonological and language skills). With regard to developmental continuity, which is the first question of interest in the present study, Scarborough's study is noteworthy. It documented that the linguistic phenotype in RD (1) appears early in development in that future RD children had less developed expressive syntax at age 2, (2) is not exclusively phonological, and (3) changes its manifestation in the preschool period.

Although the predictors of normal and extreme variations in literacy skill appear similar across these various studies, only one study directly tested the degree of similarity (Gallagher et al., 2000). Letter-name knowledge was the strongest predictor in both groups, but performance IQ (PIQ) and a language factor were predictive only in the HR group, whereas a speech factor (articulation and nonword repetition) was predictive only in the LR group. Directly testing the similarity of predictors across risk groups was one goal of the present study.

Is Familial Risk Discrete or Continuous?

If what is transmitted in dyslexic families is a single, discrete risk factor such as a major gene (e.g., Pennington et al., 1991), then nondyslexic children in these families would be expected to be similar to controls from nondyslexic families. If, on the other hand, there are multiple familial risk factors for dyslexia, then some of these would be transmitted to nondyslexic children in dyslexic families and consequently their reading development would not be identical to that of controls.

Such a result was found by Elbro et al. (1998). They found that children of RD parents, whether RD ($N = 18$) or non-RD (NRD; $N = 31$), were impaired on morpheme deletion and articulatory accuracy and efficiency, whereas those who later became RD had additional deficits in letter naming, phoneme awareness, verbal STM, and distinctness of phonological repre-

sentations. In contrast, the results of Scarborough (1989, 1990, 1991) and Gallagher et al. (2000) lend some support to the view that risk is discrete because these studies found no differences between HR NRD children and LR controls.

Existing longitudinal studies thus disagree with regard to whether family risk is discrete or continuous. Possible reasons for this disagreement are differences in diagnostic criteria for RD or differences in sample size of the HR NRD group. For instance, Scarborough had a lenient definition of RD, which would have reduced the rates of literacy problems in the HR-NRD group, which was also quite small ($N = 11$). Elbro et al. (1998), with the only results supporting continuous risk, also had the largest sample ($N = 31$) of such children.

Accurate Prediction of Later RD?

Two of the longitudinal studies just reviewed (Elbro et al., 1998; Scarborough, 1989, 1990, 1991) addressed this question and found similar results; namely that preschool measures could predict later RD with moderate accuracy (82% to 84% correct classification rate).

In sum, the issues addressed in this longitudinal study of children at family risk for RD were (1) the continuity of the phonological phenotype, (2) the similarity of predictors of reading skill, (3) the continuity of family risk, and (4) the accuracy of prediction of later RD.

METHODS

Participants

One hundred thirty-three potential participants for the study were recruited from two sources: (1) volunteers from 10 Denver area preschools, and (2) volunteers from families with a history of RD. The majority of the participants in the study came from the Denver area preschools; 10 of the HR participants were recruited from the families with a history of RD. The volunteers from these 10 families were ascertained in two ways: Their parents either were members of an organization related to learning disabilities or had an older RD child in a special education program in a local public school.

The self-selection bias inherent in a volunteer sample was expected to limit the generalizability of the results to other populations. It was important, however, to find families who were committed to completing the study. Attrition rates in longitudinal studies can be very high, and because of the nature of

the study and the population being recruited, it was felt that the stability of the sample outweighed the loss of some generalizability.

Of the 133 families who volunteered for the study, 9 were excluded, leaving 124 families. Of these nine, 2 were excluded because of confounding factors (head injury or low IQ) in the child, 5 because family risk was not parental (siblings but not parents had RD), and 2 because the RD status of the parents remained ambiguous after the procedure described later. Of the 124 children who started the study, 107 completed the entire study, and 113 children completed Time 3, the first point at which diagnoses were made. Thus, attrition was quite low. There were no significant differences between children who completed the study and those who did not in terms of age, IQ, socioeconomic status (SES), or gender ratios.

When a parent contacted project personnel regarding study participation, a brief telephone interview was conducted to determine whether there was evidence of a positive parental history of RD. Children from families in which at least one parent reported reading problems were classified as HR, and those who had no family history of reading problems were classified as LR.

To further document the telephone interview, a written Reading History Questionnaire (RHQ; Lefly & Pennington, 2000) was sent to each parent. The RHQ was a revision of a questionnaire designed by Finucci, Isaacs, Whitehouse, and Childs (1984). In two separate samples, this revised RHQ was both reliable and valid, Cronbach's $\alpha = .92$ and $.94$, test-retest reliability = $.84$ and $.87$, and valid, correlations with actual reading measures ranged from $.69$ to $.84$. If at least one parent had an RHQ score greater than or equal to $.30$ (an empirically derived cutoff based on test results in the validity study), that family was classified as HR; otherwise, the family was classified as LR. In later analyses, the RHQ variable is the higher of the two parental scores.

Eleven parents had risk classifications based on the initial interview that did not agree with the RHQ. Their ambiguous status was resolved in one of two ways: Parents were sent a more detailed RHQ to complete, or parents were interviewed in person and given a battery of reading and spelling tests. Of the 11 ambiguous cases, 9 could be resolved in this fashion; the 2 that could not be resolved were dropped from the study. This procedure for assessing parental reading problems separated the 124 families into 67 in the HR group and 57 in the LR group.

Table 1 describes the HR and LR samples at the beginning of the study. As expected, the RHQ scores of the HR parents were significantly higher than those of the LR parents, but otherwise the two groups were similar in age, gender, SES, and IQ, despite not being matched on these last three variables. In terms of ethnicity, 8% of the total sample was non-White, with similar proportions in the HR and LR samples. The 10 non-White participants included 2 Asians, 2 African Americans, and 6 Hispanics.

Hollingshead SES classifications (1975) were based on the educational level and job status of the head of household, with lower numbers indicating more status. As can be seen in Table 1, the mean SES score reflects similar middle to upper-middle-class status for both risk groups and is higher than the metro Denver average, according to the 1990 census. Although not representative of the metro Denver population, this sample is less likely to have reading problems that result from reduced language and preliteracy experiences in the home environment (Heath, 1983). Hence, it is an ideal sample in which to study the influence of genetic risk on reading development.

Generally, children began the study during the summer before they entered kindergarten and completed it the summer after second grade. At each of their four visits they were administered reading, phonological, and IQ measures (Table 2).

Table 1 Demographic Variables

Risk Group	Hollingshead SES			Age		IQ Comp ^a		Vocabulary		Block Design		RHQ	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
High	67	2.37	.9	5.4	.4	113.1	11.2	12.00	2.9	12.19	3.0	.47	.16***
Male	38	2.13	.9	5.4	.4	115.9	11.4	12.06	2.8	12.29	3.3	.49	.14
Female	29	2.67	1.0	5.3	.4	109.1	9.8	11.92	3.0	12.04	2.7	.46	.17
Low	57	2.13	.8	5.3	.3	115.5	11.5	12.72	2.9	13.15	2.4	.22	.16
Male	29	2.03	.8	5.3	.3	116.1	11.9	13.12	3.3	13.48	2.5	.22	.12
Female	28	2.22	1.0	5.3	.3	114.9	11.4	12.30	2.4	12.78	2.4	.23	.09

Note: SES = family socioeconomic status; RHQ = parental Reading History Questionnaire.

^a Estimated from the Wechsler vocabulary and block design subtests and averaged across the four time points.

*** $p < .001$.

Table 2 Tests Administered over the Longitudinal Study Time Points

Tests Administered	Time 1	Time 2	Time 3	Time 4
IQ and Achievement tests				
WPPSI, WISC-R, or WISC-III (vocabulary and block design subtests)	•	•	•	•
Letter-name knowledge	•	•	•	
Single-word reading task	•			
Woodcock-Johnson				
Letter-word identification		•	•	•
Passage comprehension			•	•
Word attack			•	•
Calculations				•
Applied problems				•
Gray Oral Reading Test			•	•
Wide Range Achievement Test				
Spelling subtest			•	•
Peabody Individual Achievement Test				
Math subtest			•	
Phonological processing				
Speech perception/repetition	•	•	•	•
Rapid serial naming				
Rapid Automatized Naming	•	•	•	•
Rapid Automatized Sequencing	•	•	•	•
Phonological awareness^a				
Roswell-Chall Test of Auditory				
Blending (Sub, Phon, S)	•	•	•	
Syllable tapping (Syl, A)	•			
Sound Categorization Test (Sub, Phon, A)				
Onset-rime (Sub, A)	•	•	•	•
Initial consonant different (Phon, A)	•	•	•	
Supply initial consonant (Phon, A)		•	•	
Strip initial consonant (Phon, A)		•	•	
Lindamood (Phon, M)		•	•	•
Pig Latin production (Phon, M)			•	•
Phoneme deletion (Phon, A)				•
Phoneme reversal (Phon, M)				•
Verbal short-term memory				
Real words	•	•	•	•
Rhyming words	•	•	•	•
Pseudo words	•	•	•	•

Note: WPPSI = Wechsler Preschool and Primary Scale of Intelligence; WISC-R = Wechsler Intelligence Scale for Children-Revised; WISC-III = Wechsler Intelligence Scale for Children-III.

^aPhonological awareness tasks varied both in size of phonological unit (syllable: Syl, subsyllable: Sub, or phoneme: Phon) that was processed and the type of processing required (analysis: A, synthesis: S, and manipulation: M).

Measures

Literacy and IQ

A single-word reading task was developed for this study and administered at the outset of Time 1 testing to identify participants who were already reading. This

task consisted of 36 words taken from basal readers, common young children's books, and the lowest levels of standardized reading tests. This task has an internal consistency reliability of .93 (Cronbach's α).

Several dimensions of literacy skill were evaluated in the study, including (uppercase) letter-name knowledge (Times 1 and 2), single-word recognition accuracy (Woodcock-Johnson, or WJ, letter-word identification subtest, at Times 2, 3, and 4), reading comprehension (WJ passage comprehension at Times 3 and 4), nonword reading accuracy (WJ word attack at Times 3 and 4), accuracy of single-word spelling (Wide Range Achievement Test, or WRAT, spelling subtest at Times 3 and 4), and speed and accuracy of reading text (Gray Oral Reading Test, or GORT, at Times 3 and 4).

Overall IQ was estimated from two Wechsler subtests, the vocabulary and block design subtests of the age-appropriate Wechsler Intelligence test—the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) at Time 1, Wechsler Intelligence Scale for Children-Revised (WISC-R) at Times 2 and 3, and the Wechsler Intelligence Scale for Children-III (WISC-III) at Time 4. Sattler (1992, p. 368) states "these two subtests have excellent reliability, correlate highly (.92) with the full scale over a wide age range, and are good measures of 'g'."

Phonological Processing

Speed perception/repetition. This task was adapted from the task used by Brady, Shankweiler, and Mann (1983) and included high-frequency words ($N = 24$) and nonwords ($N = 24$) presented in quiet and noise. Participants heard a mixed list of real words and nonwords randomly presented in either noise or quiet; their task was to repeat each word as accurately as possible. The reliability for this task was not reported by Brady et al. (1983).

Rapid serial naming. Two tasks were used in this domain: the Rapid Automatized Naming (RAN) Test (Denckla & Rudel, 1976), and Wolf's (1986) Rapid Automatized Sequencing (RAS) Test, both of which require continuous naming. The RAN consists of four different stimulus sets presented on continuous naming charts of 50 stimulus items presented in 10 rows of five items each. The four different stimulus sets are high-frequency lower case letters (a, d, o, s, p), digits (2, 4, 6, 7, 9), high frequency colors (red, yellow, green, blue, black), and line drawings of common objects (hand, chair, dog, star, ball). The stimuli in each set are presented 10 times in random sequence on a chart. The participants' task is to name all the stimuli on a chart as rapidly as possible in a left-to-right, top-to-

bottom order. Before the task begins, each participant is tested for basic recognition of each stimulus. Because many participants could not reliably identify numbers or lowercase letters at Time 1, the RAN score for Time 1 was limited to naming colors and objects. All RAN subtests were included at Times 2, 3, and 4.

The RAS consists of two different stimulus sets presented on continuous naming charts of 50 stimulus items presented in 10 rows of five items each. The two different stimulus sets are (1) colors, numbers, and lowercase letters and (2) numbers and lowercase letters. The administration of this task is identical to that of the RAN. The main dependent variable on both the RAN and RAS is the time taken to name all the stimuli.

Verbal STM. Three types of word lists are used in this task (Pennington et al., 1990), with 10 trials in each list: real words, rhyming words, and pseudowords. The latter two conditions each permit an evaluation of the use of phonological coding in STM in different ways. A child who is coding items phonologically in STM should suffer a greater performance decrement on rhyming items than a child who is not using such codes (or is using them to a lesser extent). In contrast, a child who is better at using phonological codes in STM should suffer a smaller performance decrement on pseudowords than a child who is less proficient at using phonological codes. Hence, if there are group differences on the use of phonological codes in STM, a Group \times Condition interaction should be found on this task.

Half the trials consisted of three items and half consisted of four items, at Times 1 and 2. At Times 3 and 4, half the trials consisted of four items and half consisted of five items. Participants heard each word list played on a tape recorder at a rate of one word per second. They were then asked to repeat the list in order.

Phonological awareness. The three major criteria for choosing phonological awareness tasks for this study were (1) reported high reliability and unique predictive validity in relation to later reading, (2) good coverage of both range of difficulty and types of items, and (3) developmental appropriateness to avoid floor and ceiling effects (Yopp, 1988). The tasks covered phonological units of different size: syllable, subsyllable, and phoneme; and different ways of processing the phonological units: analysis, synthesis, and manipulation (Table 2). Because research (Yopp, 1988) indicates that most tests of phonological awareness are significantly and positively correlated, it is possible to be confident that the same underlying skill is being assessed across age, although the measures themselves may differ.

Previous research (e.g., Yopp, 1988) suggests that

young children develop syllable and subsyllable awareness before they develop the awareness of individual phonemes. Skill on blending (synthesis) tasks also develops early. In contrast, tasks that require the child to segment or manipulate individual phonemes are much more difficult. The age at which various phonological awareness measures were given was based on this developmental progression, as can be seen in Table 2. Because all these tasks have been used in previous studies, they are described only briefly here.

The Roswell-Chall Test of Auditory Blending (Chall, Roswell, & Blumenthal, 1963) requires the subject to listen to the experimenter present sounds (e.g., "s-ing, a-t, de-sk") and tell what the whole word is. The syllable tapping task (Lieberman, 1973) requires the child to produce the correct number of taps for the number of syllables in a word presented by the experimenter. Bradley and Bryant's (1983) Sound Categorization Test is an oddity task. Participants are asked to identify which word does not sound like the others in a string of three or four words (three-item strings were used at Time 1, and four-item strings were used at Times 2–4). There are 30 trials total, 10 each focusing on the initial, medial, and final sounds, respectively, in consonant-vowel-consonant words. The onset-rime task (Treiman & Zukowski, 1988) asks the child to determine whether pairs of words have similar sounds. The 40 stimuli are presented randomly from each of three conditions: (1) 10 pairs with similar onsets (e.g., "sleep-slow"), (2) 10 pairs with similar rimes (e.g., "crunch-bunch"), and (3) 20 pairs with no similar sounds (e.g., "eat-sun").

The initial consonant different task (Stanovich, Cunningham, & Cramer, 1984) is similar to Bradley and Bryant's oddity task except that it focuses on the initial consonant only and the words do not rhyme (e.g., "boot, barn, buy, ear"). Thus, every item requires awareness of individual phonemes. The child is trained to listen to the first sound in a word and chooses the one in which the initial consonant is different. The supply initial consonant task (Stanovich et al., 1984) requires the child to listen to two words and to figure out the sound that is missing in the second that is present in the first (e.g., "cat-at; What is missing from 'at' that you hear in 'cat'"?). The strip initial consonant task (Stanovich et al., 1984) requires the child to segment off the first phoneme of a word presented by the examiner and tell what word is left (e.g., "Listen to the word 'feet.' If you take away the /f/ sound, what word is left?").

The Pig Latin production task (Pennington et al., 1990) requires the child to listen to a word (e.g., "pig" or "stay") and then generate the correct Pig Latin re-

sponse for that word (e.g., "igpay" or "taysay"). There were 10 trials at Time 3 and 30 trials at Time 4. Extensive training in Pig Latin production was provided before the administration of this task. The phoneme deletion task (Olson et al., 1994) is a more difficult version of the strip initial consonant task. It involves the taped auditory presentation of 40 nonwords. The child is asked to repeat the nonword individually, and is then instructed to delete a particular phoneme and to say the real word that remains after a phoneme has been deleted.

The phoneme reversal task was developed in the authors' laboratory (Pennington et al., in press). The task has both recognition and production components. To perform correctly on this task, the child has to be aware of individual phonemes in words, be able to segment them and rearrange them in an opposite order, and then produce or recognize the real word that results. The child receives careful instruction on the nature of the task using several two-phoneme words. In the production condition of the task, there are 24 two- and three-phoneme stimulus words that are reversible into high-frequency real words familiar to young children. After hearing the word, the child is instructed to "turn the sounds around," and report the word that results. In the recognition condition, the subject is presented with a stimulus word and two choices. The child is then asked to choose the real word that is the reverse of the sounds in the first word (e.g., stimulus word—"tea," choices: "eat" versus "it," where "eat" is the correct answer). The foils in the recognition condition are similar orthographically to the correct word and share all but one phoneme. The production condition always precedes the recognition condition.

Diagnostic Criteria

Participants were diagnosed as RD at two time points: before and after second grade (Times 3 and 4). Because RD can lead to serious educational and emotional distress, early identification and treatment of RD is important. Consequently, parents received feedback on the standardized test results from Times 3 and 4, both before and after second grade. It is possible, and even likely, that the recommended intervention during the year following the Time 3 testing improved some children's performances on the testing at Time 4. For this reason, children who were diagnosed as RD at either Time 3 or 4 were considered RD in the analyses that incorporate diagnosis. Nearly half (44%) of the RD group met criteria at both time points, and similar proportions met criteria at only one time point.

Children were diagnosed in the same way follow-

ing the third or fourth testing. Children who met two of three diagnostic criteria were diagnosed as RD. Three criteria were used to avoid floor effects on reading and spelling measures in children this young. All three criteria were based on a discrepancy between IQ and observed reading scores. Because there are questions about the validity of IQ discrepancy definitions of RD (Siegel, 1992; Stanovich, 1991), later testing was performed to see whether similar results would be obtained when using an age-discrepancy definition of RD. Because the overall agreement between the IQ and age-discrepant definitions was 86%, it was not expected that the results would vary much as a function of this diagnostic distinction.

The three IQ-based diagnostic criteria were (1) the Reading Quotient (RQ), (2) a regression-based IQ discrepancy score in which children's average observed reading and spelling scores were compared with their expected scores based on IQ, and (3) a raw-score reading quotient (RQ2). The RQ (Pennington et al., 1986) is the ratio of the average of the age-equivalent scores on the WRAT spelling subtest and the GORT divided by the average of chronological age, age for grade, and mental age. The RQ, therefore, divides children's observed performance on two very sensitive literacy measures by the performance that would be expected on the basis of their age, grade level, and IQ. Hence, RQs of 1.00 mean children are achieving at the expected level, whereas higher values mean they are achieving above expectation, and lower values mean they are achieving below expectation. Children with an RQ of .80 or less met this criterion for RD.

The regression-based IQ discrepancy score was derived by averaging children's standardized scores from the WJ letter-word identification, passage comprehension, and word attack subtests and the WRAT spelling subtest, and regressing this score on IQ in the LR sample. The resulting regression equation was used to identify children whose reading and spelling were 1.5 SDs below the value their IQ would predict.

The third criterion, the RQ2, used the same denominator as the RQ but used raw scores on three sensitive literacy measures in the numerator to avoid the floor effects associated with the first two criteria. Specifically, the RQ2 was a ratio based on the summed raw scores from the WJ word attack subtest, the WRAT spelling subtest, and the GORT, divided by the same denominator used in the RQ (the average of age, age for grade, and mental age). This ratio was then standardized relative to the mean and SD of the LR group, thereby creating a z score. The cutoff for RD on the RQ2 was 1.5 SDs below the LR mean.

The correlations among the three criteria at Time 3 ranged from .75 to .90, and at Time 4 they ranged from

.81 to .89. These correlations indicate that the three criteria measure very similar constructs. In addition, these scores exhibited longitudinal stability over a 1-year period. The correlation between the RQ scores at Time 3 and Time 4 was .71; the corresponding values for the regression and RQ2 scores was .81.

To recapitulate, children who met two of these three criteria at either Time 3 or 4 were diagnosed as RD. As expected, a higher proportion (22 of 64 children or 34%; 15 males, 7 females) of the HR participants were diagnosed as RD than were LR participants (3 of 49 or 6%; 2 males, 1 female). This result is consistent with the genetic research discussed earlier that has found that roughly between 30% and 50% of children of RD parents will develop RD. In this sample, having a parent with RD increased children's risk for RD 5.7 times over the risk found in children without RD parents. This magnitude of relative risk is also in the range of previous estimates (Gilger et al., 1991).

RESULTS

Before conducting the main analyses, a confirmatory factor analysis (CFA) was performed to test whether the four theoretical phonological processing constructs (i.e., latent variables) were adequately represented by the observed measures, and whether the factor structure differed between risk groups. The question of whether the factor structure is invariant

across groups is relevant to the issue of whether reading development is similar in the two groups (Question 2). Confirmatory factor analysis was performed using LISREL 8.0a for Windows by Jöreskog and Sörbom (1993). The CFA analysis tested the invariance of the factor structure cross-sectionally between risk groups at each of the four time points of the study. Two specific questions were addressed in the invariance analysis (Bollen, 1989): (1) Do the data for each group fit the same form, that is, have the same number of factors? and (2) Do the data for each group have the same structure, that is, the same parameter values? The former is the less restrictive analysis because it requires no between-group constraints other than equal number of factors. The second question is more complex and involves testing a series of nested models that address the amount of structural similarity.

A detailed presentation of the CFA results is contained in Lefly (1996). Briefly, the CFA revealed that four latent variables were present at each of the four points of the study. Further, the same observed variables at each time point represented the four latent constructs. The only exception to this was the phonological awareness construct, for which different measures were used at different ages. In addition, invariance analysis revealed that HR and LR participants had the same factor structures in terms of same number of factors (form) and same factor loadings (parameters).

Table 3 Correlations among Composites

Variables	Speech Perception (SPP)	Phonological Awareness (PA)	Verbal Short-Term Memory (VSTM)	Rapid Serial Naming (RSN)
Time 1				
SPP1				
PA1	.23*			
VSTM1	.25**	.61***		
RSN1	.21*	.33***	.49***	
Time 2				
SPP2				
PA2	.30**			
VSTM2	.31**	.60***		
RSN2	.22*	.49***	.38***	
Time 3				
SPP3				
PA3	.09			
VSTM3	.23*	.59***		
RSN3	.12	.55***	.51***	
Time 4				
SPP4				
PA4	.21*			
VSTM4	.24*	.60***		
RSN4	.13	.56***	.55***	

* $p < .05$; ** $p < .01$; *** $p < .001$.

On the basis of the latent variables from the CFA, raw-score composites were developed. These composites represent the four different correlated domains of phonological processing. Table 3 displays the correlations among these four latent variables. All latent variables were significantly correlated with each other at each time point, except speech perception at Times 3 and 4. Hence, these four latent variables each contain both unique and shared variance, which enables them to have different predictive relations to later literacy skills.

The phonological awareness composite included different combinations of variables at each time (Time 1: sound categorization, initial consonant different; Time 2: blending, supply initial consonant; Time 3: sound categorization, Pig Latin, strip initial consonant; Time 4: sound categorization, phoneme reversal, phoneme deletion). The correlations between the phonological awareness composites and the other composites ranged from .33 (rapid serial naming at Time 1) to .61 (verbal STM at Time 1).

Table 4 shows the longitudinal stability of the four latent variables. The variables ranged in stability from .42 to .71 for the phoneme awareness construct, which indicates a substantial relationship among the phonological awareness composites in spite of the fact that they were composed of different measures.

Table 4 Correlation Coefficients within Composites

Speech perception	SPP1	SPP2	SPP3	SPP4
SPP1				
SPP2	.16			
SPP3	.10	.39 **		
SPP4	.11	.29**	.19*	
Verbal short-term memory	VSTM1	VSTM2	VSTM3	VSTM4
VSTM1				
VSTM2	.65***			
VSTM3	.61***	.63***		
VSTM4	.51 ***	.53***	.73***	
Rapid serial naming	RSN1	RSN2	RSN3	RSN4
RSN1				
RSN2	.66***			
RSN3	.48***	.55***		
RSN4	.60***	.53***	.72***	
Phonological awareness	PA1	PA2	PA3	PA4
PA1				
PA2	.59***			
PA3	.55***	.60***		
PA4	.42***	.50***	.71***	

* $p < .05$; ** $p < .01$; *** $p < .001$.

The correlations among the verbal STM composites ranged from .51 to .73, and those for rapid serial naming ranged from .48 to .72. The speech perception composite, however, had much lower correlations, from .10 to .39, which suggests lower reliability of this measure, which in turn limits its value as a predictor. These correlations indicate that three of the four latent variables were stable across the years of the study.

Group Comparisons

The results that follow compared three groups: HR RD, HR NRD, and LR NRD. The LR RD participants were dropped from the following analyses because the size of the group ($n = 3$) was too small to provide meaningful comparisons. All post-hoc analyses were conducted using the Student-Newman-Keuls procedure ($\alpha = .05$).

IQ and Achievement Tests

Table 5 presents the RHQ, IQ, and achievement data for the three groups. It is noteworthy that the two HR groups did not differ on parental RHQ, with the HR NRD group in fact having a nonsignificantly higher parental RHQ score than the HR RD group. Thus, degree of parental reading problems was not confounded with diagnosis in the two HR groups.

Secondly, neither risk status nor diagnosis was confounded with IQ. As can be seen, the three groups did not differ on Time 4 IQ, just as the two risk groups did not differ on IQ (Table 1). The mean IQ in each group was above average, consistent with the demographic characteristics of these families.

The means for all three groups were at least at the population average of 100 on the three mathematics achievement tests, although by Time 4, the HR RD group was underperforming relative to the other two groups. This result is consistent with the common finding that RD is associated with some difficulty in arithmetic, presumably because of greater difficulty with reading "word" problems and memorizing math facts.

More striking group differences were found on the literacy measures. Lower scores in the HR RD group were inevitable, given that these measures were used to define RD. As can be seen, the means for the HR RD group on these reading and spelling measures were generally about 1 *SD* below the population mean and about 1.5 *SDs* below the mean of the LR RD group. The HR RD group's literacy performance therefore was significantly below both age norms and IQ expectation.

Of greater interest is the fact that the HR RD group differed significantly from the other two groups at

Table 5 IQ and Achievement Measures by Group at Times 1–4

	High-Risk RD		High-Risk NRD		Low-Risk NRD	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Parental RHQ	.45 _a	.15	.50 _a	.17	.23 _b	.11
Letter-name knowledge						
Time 1	11.7 _a	9.4	21.0 _b	6.8	21.5 _b	6.4
Time 2	22.5 _a	4.5	25.4 _b	1.8	25.6 _b	1.0
Time 3	25.3	1.0	25.9	1.0	26.0	0
WJ letter-word identification						
Time 2	84.7 _a	9.8	102.8 _b	12.3	101.7 _b	17.0
Time 3	85.5 _a	8.4	107.6 _b	13.6	115.7 _c	15.3
Time 4	90.2 _a	10.0	107.4 _b	14.7	115.0 _c	15.3
WJ word attack						
Time 3	84.5 _a	9.1	104.2 _b	13.2	108.4 _b	11.1
Time 4	88.3 _a	11.3	102.2 _b	13.2	108.2 _c	16.6
Gray Oral Reading Test						
Time 3	1.3 _a	.3	1.9 _b	.7	2.4 _c	1.1
Time 4	1.6 _a	.4	2.7 _b	1.3	3.4 _c	1.7
WRAT spelling						
Time 3	76.3 _a	9.1	90.7 _b	11.5	100.2 _c	13.3
Time 4	74.7 _a	9.9	89.3 _b	15.4	99.3 _c	13.9
IQ (Time 4)	112.1	10.3	113.1	11.1	115.6	11.7
Mathematics						
PIAT (Time 3)	104.1	8.5	108.6	10.6	110.9	11.1
WJ calculations (Time 4)	101.5 _a	15.9	111.2 _b	18.1	113.1 _b	18.5
WJ applied problems (Time 4)	110.4 _a	15.9	115.7 _b	15.4	121.2 _b	16.0

Note: Pairs with different subscript letters differ significantly ($p < .05$). All values are standard scores, except for letter-name knowledge (number of letters) and Gray Oral Reading Test (grade equivalent). RD = reading disability; RHQ = Reading History Questionnaire; WJ = Woodcock-Johnson; WRAT = Wide Range Achievement Test; PIAT = Peabody Individual Achievement Test; NRD = non-RD.

Times 1 and 2 on knowledge of letter names, a measure not used to define diagnostic groups. By Time 3, all groups were at ceiling on this measure. As will be seen, letter-name knowledge was the most powerful predictor of reading skill in the HR group, and it was the strongest predictor of diagnostic status. Therefore, across the first 2 years of the study, letter-name knowledge was still developing in all three groups, but it was developing more slowly in the HR RD group.

Despite this difference in letter-name knowledge at Time 1, the groups did not differ on entry to the study on the single-word reading task. Children who were able to read more than three words on this measure were considered to be early readers. There were only 13 early readers at Time 1, 1 in the HR RD group and 6 each in the other two groups. Neither the proportions of early readers at Time 1 nor the mean scores on this measure, $M (SD)$: .5 (1.7), 2.1 (4.5), and 3.1 (8.0) in the HR RD, HR NRD, and LR NRD groups, respectively, differed significantly among the three groups. The results of the main analyses, presented later, did not change when the 13 early readers were dropped from the sample.

Also of interest is the fact that the HR NRD group was significantly lower than the LR NRD group on most of the reading and spelling measures, despite the fact that both groups met the same criteria for being NRD. Their average difference was about .5 SD . This result supports the hypothesis that familial risk is continuous rather than discrete. As will soon be seen, the results from the phonological measures also support that hypothesis.

Phonological Processing Measures

These results may be summarized very simply. There were main effects for group, with the HR RD group always performing worst, the LR NRD group performing best, and the HR NRD group performing in between. There were also main effects for time. Because all groups showed similar rates of developmental gain across the four time points of the study, there were not significant Group \times Time interactions.

On the speech perception composite (Figure 1), there was a main effect for diagnostic group, $F(2, 101) = 5.13, p < .01$, and a main effect for time, $F(3, 303) =$

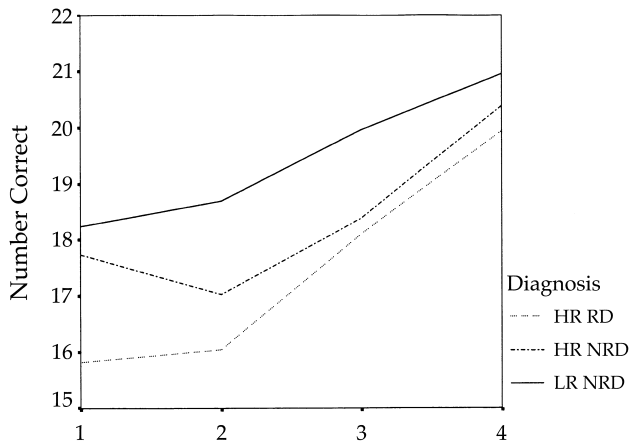


Figure 1 Speech perception by time and diagnosis. HR RD = high-risk reading disability; NRD = non-reading disability; LR = low-risk.

16.52, $p < .001$, but no Group \times Time interaction. The only significant pairwise group effect was that the HR RD children were significantly lower than the LR NRD children on speech perception at Time 2.

For the phoneme awareness composite (Figure 2), there was an overall main effect for diagnostic group, $F(2, 101) = 11.11$, $p < .001$, a main effect for time, $F(3, 303) = 28.00$, $p < .001$, and no significant interaction. Post-hoc analyses of relevant simple contrasts revealed that, in general, the HR RD group was significantly lower than both NRD groups on phoneme awareness. The sole exception was at Time 1, when they were different only from the LR NRD group. Because different measures were used at different times, the phoneme awareness composite scores did not always increase across time points.

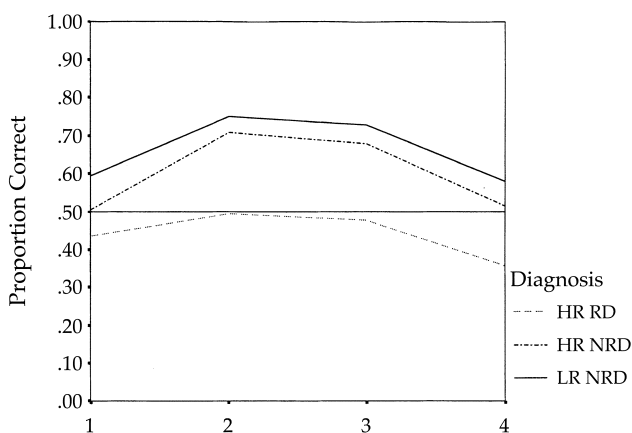


Figure 2 Phoneme awareness by time and diagnosis. HR RD = high-risk reading disability; NRD = non-reading disability; LR = low-risk.

On the verbal STM composite (Figure 3), there was an overall main effect for diagnostic group, $F(2, 101) = 7.95$, $p < .001$, a main effect for time, $F(3, 303) = 30.71$, $p < .001$, and no significant interaction. Post-hoc analyses revealed that there were no differences among the groups at Time 1. At Time 2, the HR RD group was significantly lower than either NRD group; at Time 3, both HR groups were significantly lower than the LR NRD group; and at Time 4 all three groups were significantly different from one another (HR RD $<$ HR NRD $<$ LR NRD).

Recall of real, rhyming, and pseudowords was also compared across groups to test for differences in the use of phonological coding in STM. The results are presented in Table 6. The significant group main effects have already been discussed. As expected, there were robust main effects for word type, $F_s(2, 202) = 199$ to 462, $p < .001$ at each time point, which reflects the fact that all groups performed substantially worse in the pseudoword condition at all ages and somewhat worse in the rhyming condition at most ages. The rhyming effect was not observed in any group at Time 4, which suggests that all groups "outgrew" it by age 8; it was not yet present in the HR RD group at Time 1. The key test for group differences in the use of phonological coding in STM was whether there were Group \times Word Type interactions. As can be seen in Table 6, there was a trend toward such an interaction at Times 1 and 2, and the interaction was significant at Times 3 and 4. The basis of these interactions can be examined by quantifying the rhyming and pseudoword effects as the percent decrement in word span relative to the real word condition. The rhyming effect is always greater in the LR NRD group, *mean* decrement = -7.8% , *range* = 0% to -13% , than in the HR

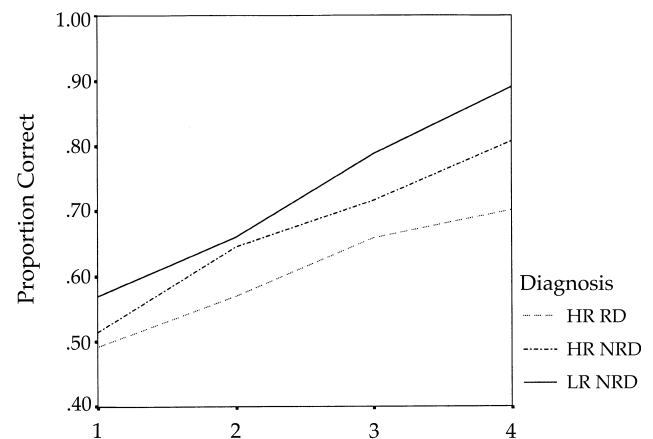


Figure 3 Verbal short-term memory by time and diagnosis. HR RD = high-risk reading disability; NRD = non-reading disability; LR = low-risk.

Table 6 Verbal STM Results by Condition, Time Points, and Diagnosis

STM Results (Proportion Correct)	High-Risk RD		High-Risk NRD		Low-Risk NRD		Group × Condition F(4, 202)
	M	SD	M	SD	M	SD	
Real1	.55	.22	.62	.20	.66	.17	
Rhym1	.58	.17	.55	.15	.59	.14	2.03 ⁺
Pseudo1	.27	.15	.30	.16	.38	.19	
Real2	.71	.13	.75	.15	.79	.14	
Rhym2	.65	.15	.69	.11	.69	.11	2.29 ⁺
Pseudo2	.27	.16	.41	.19	.41	.20	
Real3	.61	.11	.64	.09	.68	.08	
Rhym3	.58	.09	.59	.09	.63	.08	4.90 ^{**}
Pseudo3	.27	.16	.37	.14	.44	.14	
Real4	.63	.16	.73	.13	.77	.15	
Rhym4	.65	.11	.75	.09	.77	.11	2.88 [*]
Pseudo4	.27	.14	.32	.16	.44	.21	

Note: STM = short-term memory; RD = reading disability; NRD = non-RD; Real = real words; Rhym = rhyming words; Pseudo = pseudowords.
^{*}*p* < .05; ^{**}*p* < .01, ⁺*p* < .10.

RD group, mean decrement = -1.7%, range 5% to -8%, whereas the pseudoword effect was always smaller in the LR NRD group, mean decrement = -42%, range = -35% to -48%, than in the HR RD group, mean decrement = -56.5%, range -51% to -62%. In other words, as predicted, the LR NRD group used phonological codes in STM more efficiently than the HR RD group, which resulted in a rhyming penalty and a pseudoword advantage. As in other analyses, the values for the HR NRD group were in between those for the other two groups for both the rhyming effect, mean decrement = -6%, range = 3% to -11%, and the pseudoword effect, mean decrement = -48.8%, range = -%42 to -56%. In sum, both HR groups had a lower verbal STM span than the LR NRD group, and this lower span may be partly explained by less efficient use of phonological coding.

On the rapid serial naming composite (Figure 4), there was an overall main effect for diagnosis, *F*(2, 101) = 16.98, *p* < .001, and a main effect for time, *F*(3, 303) = 308.02, *p* < .001, but no significant interaction. Simple contrasts revealed that the HR RD children were significantly slower than the LR NRD children at Time 1. At Time 2, the HR RD children were significantly slower than the HR NRD and LR NRD children. At Times 3 and 4, the HR RD children were significantly slower than the HR NRD children, and the HR NRD children were significantly slower than the LR NRD children.

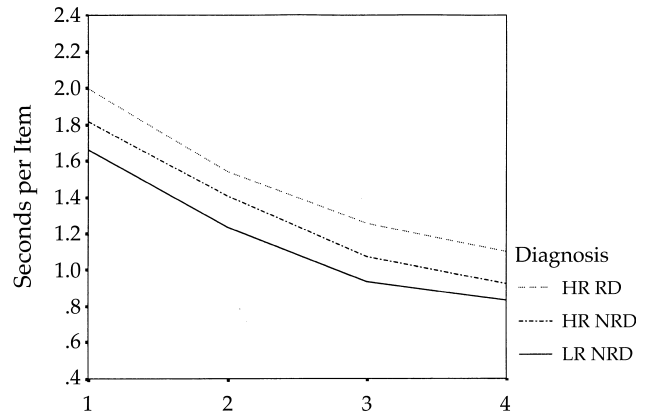


Figure 4 Rapid serial naming by time and diagnosis. HR RD = high-risk reading disability; NRD = non-risk reading disability; LR = low-risk.

Similar results for the phonological processing measures were found when RD was defined according to age-discrepant criteria or when the sample was restricted to participants who were nonreaders at Time 1.

Summary of Diagnostic Group Comparisons

To assist in the interpretation of the diagnostic contrasts, effect sizes were calculated for all the pairwise comparisons on the phonological variables. The average effect size for the difference (relative to the LR NRD group) in the HR RD group was .88; that for the HR NRD group was .37. In the HR NRD group, this difference was significant on implicit phonological processing measures (verbal STM and rapid serial naming), but not on explicit phonological processing (phonological awareness). On both literacy and phonological processing measures, therefore, the HR NRD group performed worse than the LR NRD group, which clearly supports the hypothesis that family risk is continuous rather than discrete (Question 3).

With regard to Question 1, the phonological phenotype in children who become RD is clearly broader than it is at later ages, because the HR RD group had deficits in all four aspects of phonological processing. To pursue this question further, the HR RD group at Time 4 was also compared with an RA-matched group of LR NRD participants at Time 3. It was possible to match 19 pairs on the WJ letter-word identification raw scores. Even at Time 3, these RA controls outperformed the RD participants at Time 4 on phonological measures. Specifically, the two groups differed on percent of items recalled on the verbal STM measure, RD: *M* = 68.8, *SD* = 15.6 versus RA: *M* = 78.2, *SD* = 13.8, *t* = 1.96, *p* < .05, RAN/RAS seconds

per item, RD: $M = .99$, $SD = .35$ versus RA: $M = .77$, $SD = .11$, $t = 2.64$, $p < .05$, and percent correct on phoneme awareness, RD: $M = 47.0$, $SD = 4.1$ versus RA: $M = 67.0$, $SD = 1.9$, $t = 10.81$, $p < .001$. The groups did not differ on raw scores on the speech perception measure, RD: $M = 20.0$, $SD = 3.8$ versus RA: $M = 19.7$, $SD = 3.1$. There was a trend toward a difference on the nonword reading measure, RD: $M = 5.9$, $SD = 5.8$ versus RA: $M = 8.2$, $SD = 4.6$, $t = 1.56$, $p > .10$. Thus, the RD participants at Time 4 had a broader phonological deficit than older RD participants in many previous studies, even when an RA comparison was used, but they had not yet clearly developed a nonword reading deficit relative to RA controls.

Similarity in Predictors of Normal and Extreme Variations

To address Question 2, stepwise multiple regressions in each risk group were performed to examine how well five variables (letter-name knowledge and the four phonological composites) collected at Times 1, 2, and 3 predicted reading and spelling at Time 4. The results are contained in Table 7.

These analyses test whether there are risk group differences in literacy predictors and how such predictors vary with age or literacy outcome measure. As can be seen, there are four literacy outcome measures: single word reading accuracy (WJ letter-word identification), nonword reading accuracy (WJ word attack), single-word spelling (WRAT), and fluency of reading connected text (GORT). For the LR group,

prediction did not vary much by age or literacy outcome measure. In all cases except one (predicting Time 4 spelling by using Time 1 measures), phonological awareness was the main and usually only predictor: It accounted for between 18% and 39% of the outcome variance. In contrast, for the HR group, prediction varied markedly by age but not by outcome measure. Prediction from Times 1 and 2 was dominated by letter-name knowledge but shifted by Time 3 to phonological awareness. The HR group, therefore, underwent a developmental shift by Time 3 that had mostly occurred by Time 1 in the LR group. A vestige of this developmental shift was observed in the LR group for prediction of the spelling measure. At Time 1, letter-name knowledge solely predicted Time 4 spelling outcome in both groups. By Time 2 in the LR group, the predictor of spelling outcome had shifted to phonological awareness.

The answer to the question of whether the predictors of literacy vary by risk group therefore depends on the age at which you ask the question. By Time 3, phonological awareness was similarly predictive in both risk groups. At Times 1 and 2, the predictors differed. These data are thus consistent with the view that both risk groups were traversing a similar developmental pathway at different rates. These data were also consistent with the view that letter-name knowledge develops before phonological awareness.

Prediction of Future RD

A discriminant function analysis was performed by using Time 1 variables to see how accurately fu-

Table 7 Prediction of Time 4 Literacy Skills by Risk Group

	High Risk				Low Risk			
	SW	NW	Sp.	Fl.	SW	NW	Sp.	Fl.
From Time 1								
Letter-name knowledge	.287	.177	.218	.151	.063		.176	
Phonological awareness					.258	.182		.279
From Time 2								
Letter names	.209	.211	.166	.150				
Phonological Awareness					.388	.330	.318	.303
Verbal short-term memory	.083							
Rapid serial naming			.068					
From Time 3								
Letter names								
Phonological awareness	.505	.444	.456	.206	.377	.322	.387	.259
Verbal short-term memory			.042					.062
Rapid serial naming			.041					

Note: Values are the change in R^2 . SW = Single-word recognition accuracy (Woodcock-Johnson letter-word identification); NW = nonword reading accuracy (Woodcock-Johnson word attack); Sp. = Wide Range Achievement Test spelling; Fl. = Gray Oral Reading Test.

ture diagnostic status could be predicted. Seven variables were used, all of which contributed to the resulting discriminant function. These variables and their correlations with the function were letter-name knowledge (.85), RHQ (−.27), IQ (.16), speech perception (.36), phonological awareness (.37), verbal STM (.27), and rapid serial naming (−.46). This discriminant function was calculated with a combined $\chi^2(7, N = 110) = 33.58, p < .001$. It correctly classified 69.2% of the RD sample and 76.2% of the NRD sample, with an overall classification rate of 74.5%.

The results of the discriminant function analysis indicate that letter-name knowledge and rapid serial naming of colors and objects were most important in predicting later RD. A second discriminant analysis was performed after statistically controlling for letter-name knowledge. This time, no significant discriminant function was calculated, $\chi^2(6, N = 110) = 8.514, ns$. This result indicates that no additional variance was accounted for by rapid serial naming after controlling for letter-name knowledge. This result is consistent with the results of Wagner et al. (1997), who found no predictive value for rapid serial naming after controlling for letter-name knowledge.

The classification results of the present study were also compared with those in Scarborough's (1998) meta-analysis. The sensitivity (.49 versus .78 in the meta-analysis) and specificity (.76 versus .91 in the meta-analysis) in the present study are lower, probably because the sample was weighted in favor of the HR group and because the present study lasted longer than most studies Scarborough reviewed (3 years versus 2 years). Because the HR NRD group had worse performance on phonological processing tasks, classifying them correctly was more difficult.

Determining if the accuracy of the present study's predictions would be improved if family risk status were the first step in a two-step screening procedure was of interest. The change from the previous discriminant analysis was that the sample was first divided into HR and LR groups on the basis of the results of the RHQ, and the discriminant analysis was then conducted within the HR group. Such a two-step screening process might be feasibly implemented within a school district. The first step uses family history and classifies all LR children as NRD (with a 6% false negative rate). The second step uses the more time-consuming behavioral tests in what would be a much smaller HR group. The same variables were entered (except RHQ, of course), and all of them except IQ contributed significantly. These variables and their correlations with the resulting function were letter-name knowledge (.90), phonological awareness (.33), rapid serial naming (−.31), speech perception (.31),

and verbal STM (.11). This discriminant function was, therefore, similar to the one based on the entire sample, which means that essentially the same variables that discriminated the RD group from the HR NRD group also discriminated them from the two combined NRD groups. Accuracy was improved by this two-step procedure. Sensitivity rose to .74 and specificity to .87, values similar to those in Scarborough's (1998) meta-analysis.

DISCUSSION

The goal of this project was to use the natural experiment provided by the familial transmission of dyslexia to answer questions about early reading development in this condition *before* it is diagnosable. Because genetic risk factors are transmitted randomly within families, the naturally occurring manipulation of genetic risk comes close to the random assignment required of a true experiment. The present study used this "manipulation" twice, across two generations. First, otherwise similar parents were divided into those with and without a history of dyslexia. Then otherwise similar children of parents with a dyslexic history were divided into those who later did or did not become dyslexic. If perfect knowledge of the genetic and environmental risk factors for dyslexia were available, then these two manipulations would be much cleaner. They were nevertheless clean enough to answer the questions posed in the present study about early reading development in dyslexia.

Although the design of the present study did not make possible the separation of genetic and environmental risk factors in the participating families, the demographic characteristics of the sample made it less likely that the literacy problems found in the HR group were due to environmental factors. Other research has found that the heritability of dyslexia or RD varies linearly as a function of a child's IQ. In RD children with an IQ less than 100, the heritability of RD is .43 ($\pm .09$), whereas in children with an IQ of 100 or greater, the heritability of RD is .72 ($\pm .13$), a significant difference (Wadsworth, Olson, Pennington, & DeFries, 2000). Because the mean IQ in our RD group was 112, the heritability of their RD was likely higher than .72, which means that most of the risk factors for RD in this group were genetic. Moreover, these risk factors appear to be fairly specific to phonological and literacy skills because there were no IQ differences among the groups.

The questions addressed in this study concerned (1) continuity in the phonological phenotype, (2) similarity of literacy predictors between risk groups, (3) conti-

nity of familial risk, and (4) accuracy of prediction of later RD. The answers to the first three questions are discussed later. With regard to the fourth question, later RD can be predicted with moderate accuracy at age 5 years, and the strongest predictor is letter-name knowledge. The precision of the prediction is not precise enough for diagnosis; however, it may be adequate for screening purposes and early intervention. Although this was a volunteer sample of higher-than-average SES, the results of the present study are convergent with the results of studies using other samples.

A Broader Early Phenotype

The finding in the present study of a broader phonological phenotype in 5-year-old children who later became RD is consistent with the results of all three other longitudinal studies of children at family risk discussed earlier (Elbro et al., 1998; Gallagher et al., 2000; Scarborough, 1989, 1990, 1991). This result therefore appears to be robust across the countries and languages represented by these four studies. Moreover, this broader phenotype persisted until Time 4 and was maintained when a more stringent RA comparison group was used. This broader phenotype was also found when an age-discrepancy definition of RD was used. This finding of a broader phonological phenotype contrasts clearly with results with older RD children, adolescents (Pennington et al., *in press*), and adults (Pennington et al., 1990) previously obtained using similar diagnostic procedures and similar measures of phonological processing. In those studies, the main phonological deficit found in the RA comparisons was in phoneme awareness, and few differences were found for the other three domains of phonological processing, even in the CA comparisons. On the basis of these cross-sectional comparisons, there does appear to be a clear narrowing in the breadth of the phonological phenotype in RD from early school age to later ages.

It is intriguing that evidence from the longest longitudinal study (Scarborough, 1989, 1990, 1991) of children at familial risk documents an earlier developmental shift in the phenotype of RD, from a deficit in expressive syntax to a phonological deficit. This result suggests that the genes that influence RD may transiently affect other aspects of language development, in addition to their well-documented effects on phonological development.

That these effects on phonological development have clinical consequences outside the domain of literacy is documented by the well-known associa-

tion between RD and earlier problems with expressive speech. Both retrospective (Hallgren, 1950; Rutter & Yule, 1975) and prospective (A. M. Gallagher, personal communication, December, 1996; H. S. Scarborough, personal communication, December, 1996) studies found that about 25% of RD children had clinically significant speech production problems at preschool ages, which is about two to three times the rate observed in controls. In the current sample, parents expressed concern about speech development in 28% of the HR sample as opposed to 12.5% of the LR sample, consistent with these other studies. It is clear, therefore, that a minority subset of future RD children have a clinically diagnosable speech disorder, which suggests an etiological overlap between these two disorders. Further research is needed to determine the nature of this etiological overlap.

Similarity in Predictors

The present study's finding that the predictors of reading and spelling skill were somewhat different in the two risk groups is consistent with the findings of Gallagher et al. (2000), but it is not clear that these results reflect a significant discontinuity between the two groups in the processes leading to literacy. Instead, it may be that both groups are following a similar developmental pathway but are at different developmental points on that pathway. At age 5 years, the HR group is at an earlier point in the development of letter-name knowledge than the LR group (Table 5), and consequently there is more variance in their letter-name knowledge than in the LR group's knowledge. This greater variance could produce a stronger predictive relation. The present study's results were also consistent with the view that letter-name knowledge precedes the development of phoneme awareness. The difference, therefore, in literary predictors at age 5 in the two risk groups may simply reflect the fact that they are at different stages on the same developmental trajectory.

Further evidence for similarity between the two risk groups in the developmental processes leading to literacy is provided by (1) the similar results for the two risk groups in the confirmatory factor analysis discussed earlier, and (2) a comparison of our results with Scarborough's (1998) meta-analysis of longitudinal studies of literacy development. In both the present study and the meta-analysis, letter-name knowledge was the strongest individual predictor, IQ was moderately predictive, and phoneme awareness and rapid serial naming were more predictive than either speech perception or verbal STM.

Continuity of Family Risk

The present study's finding that family risk is continuous rather than discrete agrees with the results of Elbro et al. (1998), but not those of Gallagher et al. (2000) or Scarborough (1989, 1990, 1991). The null results in the latter two studies could, however, be a function of the small sample sizes in the HR NRD groups or a function of different diagnostic cutoffs. It is unlikely that the present study's finding of literacy and phonological differences in the HR NRD group is a function of cutoffs for either risk status or RD. Inspection of the *SDs* in Table 5 and those for the four phonological composite variables reveals that they are very similar across the three diagnostic groups. If the positive findings in the HR NRD group were due to a few outliers, then the variance should be greater in that group. Instead, examination of the distributions indicated a downward shift of the whole HR NRD distribution relative to the LR NRD group.

The hypothesis that family risk is continuous is also consistent with recent results from genetic studies of RD, which suggest that multiple genes influence RD (Grigorenko et al. 1997; Smith, Pennington, Kimberling, & Ing, 1990) and that those genes are best conceptualized as quantitative trait loci, or QTLs (Cardon et al. 1994; Pennington & Gilger, 1996). QTLs (Plomin & Rutter, 1998) are genes that act to change the value of a continuous phenotypic trait, such as reading or height. In the case of an extreme phenotype like RD, each QTL acts to increase or decrease risk, but it is unlikely that any given QTL is necessary or sufficient to produce the extreme phenotype by itself. Thus, it appears likely that multiple QTLs are responsible for the moderate heritability of RD (DeFries et al. 1987). Multiple QTLs plus familial environmental risk factors would lead to continuous rather than discrete familial risk.

Molecular Measures and Future Risk Studies

The approximate location of one QTL influencing RD (on the short arm of Chromosome 6) has been replicated across three laboratories and five samples (Cardon et al., 1994; Fisher et al. 1999; Gayan et al. 1999; Grigorenko et al., 1997), although a fourth laboratory using a different phenotype definition did not replicate this result (Field & Kaplan, 1998).

If this QTL on Chromosome 6, as well as other QTLs influencing RD, can be identified and sequenced, then it might be possible to identify precisely which children of RD parents have inherited risk or protective alleles of these QTLs. Prospective studies of such children could enable us to study the development of RD

from infancy to school age. In addition, such studies could tell us whether these QTLs also cause the speech problems associated with RD and the earlier, transient problem in expressive syntax found by Scarborough (1990). Obviously, not all the children in the present study's HR group are at equal risk for RD, and some may not be at risk at all. If researchers could better characterize genetic risk, considerably more could be learned about environmental influences on RD. Some of these environmental influences could in turn be targets for early intervention.

As Plomin and Rutter (1998) have argued, therefore, incorporating molecular measures of risk into longitudinal studies has the potential to address a number of fundamental questions about the continuity of the processes involved in the development of disorders. Even with just familial measures of risk, progress toward answering these questions can be made, as illustrated in the present study.

ACKNOWLEDGMENTS

This research was supported by four National Institutes of Health grants: MH38820 (MERIT Award), MH00419 (Research Scientist Award), HD27802 (The Colorado Learning Disabilities Research Center), and HD04024 (The Colorado Mental Retardation Research Center), and by grants from the March of Dimes Foundation and the Orton Dyslexia Society.

ADDRESSES AND AFFILIATIONS

Corresponding author: Bruce F. Pennington, Department of Psychology, University of Denver, 2155 S. Race St., Denver, CO 80208; e-mail: bpenning@nova.psy.du.edu. Dianne L. Lefly is with the Denver Public Schools.

REFERENCES

- Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 38*, 446–462.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature, 271*, 746–747.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*, 419–421.
- Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345–367.

- Brady, S. A. (1997). Ability to encode phonological representations: An underlying difficulty of poor readers. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia* (pp. 21–46). Mahwah, NJ: Erlbaum.
- Bruck, M. (1993). Word recognition and component phonological processing skills of adults with childhood diagnosis of dyslexia. [Special issue: Phonological processes and learning disability.] *Developmental Review*, 13, 258–268.
- Bryant, P. E., & Bradley, L. (1985). *Children's reading problems: Psychology and education*. New York: Blackwell.
- Cardon, L. R., DeFries, J. C., Fulker, D. W., Kimberling, W. J., Pennington, B. F., & Smith, S. D. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science*, 266, 276–279.
- Chall, J., Roswell, F. G., & Blumenthal, S. H. (1963). Auditory blending ability: A factor in success in beginning reading. *Reading Teacher*, 17, 113–118.
- DeFries, J. C., Fulker, D. W., & LaBuda, M. C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature*, 329, 537–539.
- Denckla, M. B., & Rudel, R. G. (1976). Rapid "automatized" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14, 471–479.
- Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*, 20, 163–179.
- Elbro, C., Borstrom, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly*, 33, 36–60.
- Field, L. L., & Kaplan, B. J. (1998). Absence of linkage of phonological coding dyslexia to chromosome 6p23–p21.3 in a large family data set. *American Journal of Human Genetics*, 63, 1448–1456.
- Finucci, J. M., Whitehouse, C. C., Isaacs, S. D., & Childs, B. (1984). Derivation and validation of a quantitative definition of specific reading disability for adults. *Developmental Medicine and Child Neurology*, 26, 143–153.
- Fisher, S. E., Marlow, A. J., Lamb, J. M. E., Williams, D. F., Richardson, A. J., Weeks, D. E., Stein, J. F., & Monaco, A. P. (1999). A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *American Journal of Human Genetics*, 64, 146–156.
- Gallagher, A. M., Frith, U., & Snowling, M. J. (2000). Precursors of literacy delay among children at genetic risk of dyslexia. *Journal of Child Psychology and Psychiatry*, 41, 202–213.
- Gayán, J., Smith, S. D., Cherny, S. S., Cardon, L. R., Fulker, D. W., Brower, A. M., Olson, R. K., Pennington, B. F., & DeFries, J. C. (1999). Quantitative-trait locus for specific language and reading deficits on chromosome 6p. *American Journal of Human Genetics*, 64, 157–164.
- Gilger, J. W., Pennington, B. F., & DeFries, J. C. (1991). Risk for reading disability as a function of parental history in three family studies. *Reading and Writing: An Interdisciplinary Journal*, 3, 299–313.
- Grigorenko, E. L., Wood, F. B., Meyer, M. S., Hart, L. A., Speed, W. C., Shuster, A., & Pauls, D. L. (1997). Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *American Journal of Human Genetics*, 60, 27–39.
- Hallgren, B. (1950). Specific dyslexia: A clinical and genetic study. *Acta Psychiatrica Neurological Scandinavica*, 65 (Suppl.), 1–287.
- Heath, S. B. (1983). *Ways with words*. Cambridge, U.K.: Cambridge University Press.
- Hollingshead, A. B. (1975). *Four factor index of social status*. New Haven, CT: Yale University.
- Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal*, 8, 217–234.
- Jöreskog, K. G., & Sörbom, D. (1993). *LISREL VIII: User's Reference Guide*. Chicago: Scientific Software International.
- Jorm, A. F., Shore, D. L., MacLean, R., & Matthews, R. (1984). Phonological confusability in short-term memory for sentences as a predictor of reading ability. *British Journal of Psychology*, 75, 393–400.
- Lefly, D. (1996). *Risk status and phonological processing*. Unpublished doctoral dissertation. University of Denver, Denver, Colorado.
- Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, 33, 286–296.
- Lieberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society*, 23, 65–77.
- Lieberman, I. Y., Rubin, H., Duques, S., & Carlise, J. (1985). Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. In D. Gray & J. Kavanaugh (Eds.), *Biobehavioral measures of dyslexia* (pp. 163–176). Parkton, MD: York.
- Mann, V. A., & Lieberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? *Journal of Learning Disabilities*, 17, 592–598.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323–331.
- Olson, R. K., Forsberg, H., & Wise, B. (1994). Genes, environment, and the development of orthographic skills. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge I: Theoretical and developmental issues* (pp. 27–71). Dordrecht, the Netherlands: Kluwer.
- Olson, R., Wise, B., Conners, F., & Rack, J. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities*, 22, 339–348.
- Pennington, B. F., Cardoso-Martins, C., Green, P., & Lefly, D. L. (in press). Comparing the phonological and double deficit hypothesis for developmental dyslexia. *Reading and Writing*.
- Pennington, B. F., & Gilger, J. W. (1996). How is dyslexia transmitted? In C. H. Chase, G. D. Rosen, & G. F. Sherman (Eds.), *Developmental dyslexia: Neural, cognitive, and genetic mechanisms* (pp. 41–61). Parkton, MD: York.
- Pennington, B. F., Gilger, J. W., Pauls, D., Smith, S. A., Smith,

- S. D., & DeFries, J. C. (1991). Evidence for major gene transmission of developmental dyslexia. *Journal of the American Medical Association*, *266*, 1527–1534.
- Pennington, B. F., McCabe, L. L., Smith, S. D., Lefly, D. L., Bookman, M. O., Kimberling, W. J., & Lubs, H. A. (1986). Spelling errors in adults with a form of familial dyslexia. *Child Development*, *57*, 1001–1013.
- Pennington, B. F., Van Orden, G. C., Smith, S. D., Green, P. A., & Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development*, *61*, 1753–1778.
- Plomin, R., & Rutter, M. (1998). Child development, molecular genetics, and what to do with genes once they are found. *Child Development*, *69*, 1223–1242.
- Pratt, A. C., & Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *Journal of Educational Psychology*, *80*, 319–323.
- Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*, 28–53.
- Read, C., Zhang, Y. F., Nie, H. Y., & Ding, B. Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, *24*, 31–44.
- Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry*, *16*, 181–197.
- Sattler, J. M. (1992). *Assessment of children. The WPPSI-R and WISC-III supplement* (3rd ed.). San Diego, CA: Sattler.
- Scarborough, H. S. (1989). Prediction of reading disability from familial and individual differences. *Journal of Educational Psychology*, *81*, 101–108.
- Scarborough, H. S. (1990). Very early language deficits in dyslexic children. *Child Development*, *61*, 1728–1743.
- Scarborough, H. S. (1991). Early syntactic development of dyslexic children. *Annals of Dyslexia*, *41*, 207–220.
- Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In A. C. P. Accardo & B. Shapiro (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York.
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.
- Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, *25*, 618–629.
- Smith, S. D., Pennington, B. F., Kimberling, W. J., & Ing, P. S. (1990). Familial dyslexia: Use of genetic linkage data to define subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29*, 204–213.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. E. (1991). Conceptual and empirical problems with discrepancy definitions of reading disability. *Learning Disability Quarterly*, *14*, 269–280.
- Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, *38*, 175–190.
- Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language*, *27*, 466–477.
- Wadsworth, S. J., Olson, R. K., Pennington, B. F., & De Fries, J. C. (2000). Differential genetic etiology of reading disability as a function of IQ. *Journal of Learning Disabilities*, *33*, 192–199.
- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–212.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*, 468–479.
- Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language*, *27*, 360–379.
- Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly*, *26*, 123–141.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, *23*, 159–177.