Stanovich, K. E. 1991. Discrepancy definitions of reading disability: Has intelligence lead us astray? *Reading Research Quarterly* 26:7–29.

Stanovich, K. E., and Siegel, L. S. 1994. Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology* 56:24–53.

Stringer, R., and Stanovich, K. E. 1998. On the possibility of cerebellar involvement in reading disability. Paper read at the 4th conference of the SSSR, April 1998, San Diego.

Tønnessen, F. E. 1997. How can we best define 'dyslexia'? *Dyslexia* 3:78–92.

Van Daal, V. H. P. In preparation. Learning to read in a foreign language.

Van Daal, V. H. P., and Van der Leij, A. In preparation. Large-scale screening and diagnosis of dyslexia in secondary education.

Van den Bos, K. P. 1996. *BELL96*. Groningen University (Netherlands): Department of Special Education.

Van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., and De Vries, J. 1994. *De Klepel. Vorm A en B.* Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling [The Klepel. Form A and B. A test of reading pseudowords]. Nijmegen (Netherlands): Berkhout.

Van der Leij, A. 1987. Special education in the Netherlands. In *Encyclopedia of Special Education, Vol. 2*. C. R. Reynolds and L. Mann, eds. New York: Wiley & Sons.

Van der Leij, A., and Van Daal, V. H. P. 1999. Automaticity, automatization and dyslexia. In *Dyslexia: Advances in Theory and Practice*. I. Austad, I. Lundberg, and F. E. Tønnessen, eds. Dordrecht, Netherlands: Kluwer Academic Publishers.

Van der Leij, A., and Van Daal, V. H. P. In press. Automatization aspects of dyslexia: Speed limitation in word identification, sensitivity to increasing task demands, and orthographic compensation. *Journal of Learning Disabilities*.

Wimmer, H., Mayringer, H., and Landerl, K. 1998. Poor reading: A deficit in skill automatization or a phonological deficit? *Scientific Studies of Reading* 2:321–40.

Wimmer, H., Mayringer, H., and Raberger, Th. 1999. Reading and dual-task balancing: Evidence against the automatization deficit explanation of developmental dyslexia. *Journal of Learning Disabilities* 32:473–78.

Wolf, M., Bowers, P. G., and Biddle, K. in press. Naming-speed process, timing, and reading: A conceptual review. *Journal of Learning Disabilities*.

Yap, R. L., and Van der Leij, A. 1993. Word processing in dyslexics: An automatic decoding deficit? *Reading and Writing: An Interdisciplinary Journal* 5:261–79.

Yap, R. L., and Van der Leij, A. 1994. Word processing in dyslexics: An automatic decoding deficit? In *Dyslexia in Children. Multidisciplinary Perspectives*. A. J. Fawcett and R. I. Nicolson, eds. New York: Harvester Wheatsheaf.

Development of Dyslexic Subgroups: A One-Year Follow Up

Franklin R. Manis
Mark S. Seidenberg
Lynne Stallings
Marc Joanisse
Caroline Bailey
Laurie Freedman
Suzanne Curtin

University of Southern California
Los Angeles, California

Patricia Keating

University of California
Los Angeles, California

There is a consensus that dyslexia is on a continuum with normal reading skill and that dyslexics fall at the low end of the normal range in phonological skills. However, there is still substantial variability in phonological skill among dyslexic children. Recent studies have focused on the high end of the continuum of phonological skills in dyslexics, identifying a "surface" dyslexic, or "delayed" profile in which phonological skills are not out of line with other aspects of word recognition. The present study extended this work to a longitudinal context, and explored differences among subgroups of dyslexics on a battery of component reading skills. Third grade dyslexics (n = 72) were classified into two subgroups, phonological dyslexics and delayed dyslexics, based on comparisons to younger normal readers at the

Annals of Dyslexia, Vol. 49, 1999

ISSN 0736-9387

same reading level (RL group). The children were tested at two points (in third and fourth grade). The results revealed that the classification of dyslexics produced reliable, stable, and valid groups. About 82 percent of the children remained in the same subgroup category when retested a year later. Phonological dyslexics were lower in phoneme awareness and expressive language. Delayed dyslexics tended to be slower at processing printed letters and words but not at rapid automatic naming of letters, and relied more heavily on phonological recoding in reading for meaning than did phonological dyslexics. A subset of the delayed dyslexics with the traditional "surface dyslexic" pattern (relatively high pseudoword and low exception word reading) was also identified. The surface subgroup resembled the RL group on most measures and was not very stable over one year. The results are discussed in light of current models of dyslexia and recent subgrouping schemes, including the Double-Deficit Hypothesis.

DEVELOPMENT OF DYSLEXIC SUBGROUPS: A ONE-YEAR FOLLOW-UP STUDY

While there is a developing consensus among researchers that dyslexia represents the lower end of a normal distribution of reading skills (e.g., Shaywitz, Escobar, Shaywitz, Fletcher, and Makuch 1992), there is considerable interest in the question of how reading disabled individuals differ among themselves. Differences may occur in underlying cognitive skills as well as in patterns of reading development. Numerous attempts to identify subgroups of dyslexia have been made over the past two decades. This work is important, both to an understanding of the causes of reading failure, and to its prognosis and treatment. Earlier subtyping efforts utilized multivariate statistical methods to sort profiles of performance on large batteries of neuropsychological and educational tasks. This work was disappointing as it failed to reach consensus on a small number of reliable subcategories of dyslexia, and it was not grounded in a theory of how various neuropsychological tasks were related to reading (Fletcher and Morris 1986; Hooper and Willis 1989; Lyon 1985; Satz and Morris 1981; Stanovich 1991). More recent studies have used classification systems that are consistent with current views of the cognitive bases of dyslexia (Castles and Coltheart, 1993; Manis, Seidenberg, Doi, McBride-Chang, and Peterson 1996; Morris, Stuebing, Fletcher, Shaywitz, S. E., Shankweiler, Katz, Francis, and Shaywitz, B. A. 1998; Wolf and Bowers in press). In the present paper, we apply a developmental, connectionist model of reading (Harm and Seidenberg 1999) to the issue of

dyslexic subgroups. An important advantage of connectionist models is that the effects of hypothesized cognitive deficits on the reading system can be simulated, leading to interesting proposals about the nature of dyslexic reading deficits.

Three main approaches to dyslexic subgroups have been utilized in recent years. One approach relies on the phonological core hypothesis of dyslexia. According to this view, dyslexic children have poorly developed representations of words and their constituent sounds in memory. Evidence for the phonological core hypothesis is provided by repeated observations that dyslexic children have deficits on nonreading tasks involving the processing of phonological information including measures of phoneme awareness, verbal short-term memory, and rapid naming of symbols (Lyon 1995; Shankweiler, Crain, Katz, Fowler, Liberman, A. M., Brady, Thornton, Lundquist, Dreyer, Fletcher, Stuebing, Shaywitz, S.E., and Shaywitz, B. A. 1995; Stanovich 1988; Wagner and Torgesen 1987). Phonological deficits appear to have a strong genetic basis as well (Castles, Datta, Gayan, and Olson 1999; Olson, Wise, Conners, Rack, and Fulker 1989). These difficulties are thought to have widespread effects on learning to read words and may be manifested as deficits in orthographic knowledge of words at later stages of reading (Ehri 1992; Ehri and Saltmarsh 1996).

Morris et al. (1998) exemplified the phonological core framework in their cluster analytic study of poor reader subgroups. They identified seven subgroups of poor readers and two subgroups of good readers. Two poor reader subgroups were marked by a severe, global deficit across many, primarily language-based, tasks. Five subgroups had more specific deficits involving combinations of four variables: phonological awareness, verbal short-term memory, visual-spatial ability and rate of processing (including rapid automatic naming, or RAN tasks). Phonological deficits were found in four of the five specific subgroups, underlining the importance of phonological skills in reading disability. The fifth subgroup was a rate-deficient group that tended to have reading deficits on measures of reading rate but not accuracy.

In a second approach to subgrouping poor readers, Bowers and Wolf (Bowers and Wolf 1993; Wolf and Bowers in press) have argued that the rate of naming familiar symbols is a source of variation in reading ability that is partially independent of phonological skill. The hypothesis is supported by studies showing that naming speed accounts for reliable variance in reading that is not accounted for by IQ, verbal memory, and phonological awareness (e.g., Badian 1997; Bowers and Wolf 1993; Felton and Brown 1990; Manis, Seidenberg, and Doi 1999). The hypothesis is also supported by the identification of a rate-deficient subgroup in the Morris et al. (1998) cluster analytic study. Bowers and Wolf (Bowers 1995; Wolf and Bowers in press) showed that a deficit in phonological awareness and rapid naming was associated with more severe reading deficits than either deficit alone. Children with isolated rate deficits appeared to have difficulties with fluency and not accuracy of reading words (Bowers 1995; Morris et al. 1998).

Phonological deficits have also figured prominently in a third line of theoretically driven subtyping research that focuses on the processes by which words are recognized (Castles and Coltheart 1993; Castles et al. 1999; Manis et al. 1996; Murphy and Pollatsek 1994; Stanovich, Siegel, and Gottardo 1997). These studies distinguish between two main subgroups of disabled readers. Phonological dyslexics have poor phonological decoding skills (as on pseudoword reading tasks) relative to their word recognition skill (measured typically with an exception word reading task), whereas surface dyslexics show the opposite pattern (poor exception word reading relative to pseudoword pronunciation). In the dual-route model of reading used in some of these studies (q.v., Coltheart, Curtis, Atkins, and Haller 1993), phonological dyslexia is viewed as a deficit in sublexical aspects of reading, and surface dyslexia as a deficit in an independent reading mechanism operating at the lexical (whole word) level . Beginning with a series of case studies (e.g., Coltheart, Masterson, Byng, Prior, and Riddoch 1983; Temple and Marshall 1983) and continuing in several studies with large samples (e.g., Castles and Coltheart 1993; Manis et al. 1996; Stanovich et al. 1997), phonological and surface dyslexia have been shown to be valid and reliable classifications of dyslexic children. For example, phonological and surface dyslexics have been shown to differ in theoretically expected directions on tests of phoneme awareness, verbal short-term memory and orthographic knowledge, as well as in the pattern of their reading errors on words (Castles et al. 1999; Manis et al. 1996; Murphy and Pollatsek 1994; Stanovich et al. 1997).

Two interesting findings have emerged in recent studies of surface and phonological dyslexia. First, it is apparent that most dyslexics are "mixed"; that is, they have deficits in both exception word reading and nonsense word reading. Pure subgroups (children who are normal on one task and deficient on the other) are rare, and as it turns out, appear to have milder reading difficulties (Harm and Seidenberg 1999; Manis et al. 1996; Stanovich et al. 1997). A second finding is that only one of the subgroups appears to be developmentally distinctive. Surface dyslexics do not appear to differ from younger normal readers matched on overall word reading skill on a wide variety of measures, including nonword reading, phonological awareness, verbal short-term memory, orthographic skill and pattern of word mispronunciations (Manis et al. 1996; Stanovich et al. 1997). In contrast, phonological dyslexics are low on several tests of phonological skill, while scoring within or above the range of younger normal readers on orthographic and exception word tasks. Manis et al. (1996) hypothesized that surface dyslexics had delayed, but otherwise normal, word reading development, whereas phonological dyslexics deviated from the normal pattern. Longitudinal data are necessary to test this prediction. Stanovich et al. (1997) suggested that surface dyslexics might have mild phonological difficulties in combination with external environmental factors such as poor learning opportunities or low print exposure.

Castles and Coltheart (1993) interpreted the phonological and surface dyslexic profiles within the dual-route model of reading (e.g., Carr and Pollatsek 1985; Coltheart 1978; Coltheart, Curtis, Atkins, and Haller 1993). In this framework, phonological dyslexics have a deficit in a sublexical procedure that utilizes grapheme-phoneme conversion rules, and surface dyslexics are deficient in a lexical, or whole word recognition procedure by which a reader accesses a storehouse of information about whole word pronunciations.

An alternative to the dual-route model was presented by Harm and Seidenberg (1999). They proposed a comprehensive, developmental account of dyslexic subgroups using a connectionist model originally developed by Seidenberg and McClelland (1989; see also Plaut, McClelland, Seidenberg, and Patterson 1996). In this type of model, the reader gradually forms and strengthens associations between three types of units—orthographic units (which can be instantiated at the single-letter or multi-letter levels), phonological units, and semantic units—within a connectionist network that includes intervening layers of "hidden" units. The orthographic and phonological part of the model has been implemented as a computer simulation which first learns phonology, then is exposed to learning trials with a corpus of 3,123 printed monosyllabic words (Harm and Seidenberg 1999). The model learns to read both regular and exception words, and also acquires the ability to make generalizations to unfamiliar words and pseudowords.

Harm and Seidenberg (1999) and Manis et al. (1996) argued that connectionist models account for two phenomena that the dual-route model does not. First, most dyslexics have deficits on both pseudowords and real words, and differ primarily in the degree of such deficits rather than in some absolute sense; indeed many dyslexics have a "mixed" profile. The connectionist model produces a mixed pattern if the phonological units are damaged, simulating problems with the quality of phonological representations in memory. A slight amount of damage affects primarily pseudoword reading, but greater damage has effects on both pseudowords and exception words (Harm and Seidenberg 1999). The dual-route model accounts for pure cases (who have one deficit but not the other) but not for mixed cases. Secondly, the dual-route model, with its emphasis on grapheme-phoneme conversion, does not take the strong evidence for phonological deficits on nonreading tasks into account, whereas connectionist simulations have the ability to examine explicitly the effects of poor phonological representations on reading.

Harm and Seidenberg also simulated the effects of a variety of deficits in nonphonological processes within the model. What is interesting about these simulation studies is that several deficits tended to produce the delayed but normal pattern characteristic of surface dyslexia. Thus, deficits in the amount of hidden unit resources, deficits in overall learning rate, visual-orthographic deficits, and lowered print exposure all had the same effect, producing a general retardation in the development of orthographic to phonological associations that made the model look like models in the early phases of learning. During early phases, the models generally have learned regular words but have not mastered many of the exception words, much like the surface dyslexic pattern.

The Harm and Seidenberg model leads to an interesting set of predictions about the kinds of reading deficits that should be seen. Dyslexic children should fall into two broad categories: those with phonological deficits and those with delayed but normal reading acquisition. Surface dyslexics may be identifiable but would be a subset of individuals in the general delay category. The delayed group most likely has a heterogeneous set of causes and might be broken down further into subgroups associated with these causes. However, it is difficult to make inferences about causes from concurrent data on test batteries, so caution must be exercised.

How do the two subtyping frameworks reviewed earlier intersect with the connectionist model? Children with deficits in phonological awareness and verbal short-term memory identified by Morris et al. (1998) should tend to fall into the phonological dyslexic category, whereas children with deficits in other types of skills (e.g., visual-spatial ability or low overall cognitive ability) should tend to fall either in the delayed category or the normal reader category. Deficits in processing rate (or in RAN performance) might be found in both the phonological and delayed subgroups, as many children with phonological deficits may have double deficits, and some children with general delays might have mild phonological deficits in combination with slow RAN times. Slow symbol naming may be related to exception word learning, as both tasks involve accessing arbitrary item-specific information from the point of view of the connectionist model (see Manis et al. 1999 for a more extended discussion of RAN within a connectionist model). However, exception word reading deficits in the absence of serious phonological difficulties would tend to place an individual in the delayed dyslexic category. Children with RAN-only deficits might be deficient in reading fluency rather than accuracy, and, therefore, not be identified as poor readers using accuracy-based reading tests (Morris et al. 1998; Wolf and Bowers in press).

In sum, the recent literature on dyslexic subgroups seems to converge on two points: (1) phonological deficits are prominent among dyslexic children and appear to be the most valid classification variables; and (2) there are dyslexic children who do not appear to have core phonological deficits. The latter subgroup seems to fit a developmentally normal but delayed pattern. However, the nature of the cognitive deficit in this group is not clear.

The present study was designed to extend the recent work on dyslexic subgroups by assessing a younger sample of dyslexics that could be followed longitudinally. Dyslexics were divided into two subgroups based on comparisons to younger normal readers matched on word recognition skill (referred to hereafter as the reading level, or RL, comparison group): a phonological subgroup (deficient on phonological awareness or pseudoword reading); and a delayed subgroup (equally delayed in word recognition but not deficient on either of the phonological tasks).

Defining dyslexic subgroups based on a RL comparison group is a straightforward application of Harm and Seidenberg's (1999) simulation studies, as well as previous studies in which dyslexics emerged as either phonologically impaired or delayed in reading. Subgroups defined in this way will not necessarily resemble pure phonological and surface dyslexics; some of the phonological cases defined here would be pure phonological cases, and some of the delayed cases would be expected to have a surface dyslexic profile. Phonological dyslexics in the current scheme have the lowest phonological skill in the sample, and according to the Harm and Seidenberg model, would be expected to have both nonword and exception word reading difficulties, as well as deficits in orthographic knowledge. Delayed dyslexics do not have serious phonological deficits but read at about the same grade level. Therefore, they can be expected to be comparable in exception word and orthographic skills to the phonological group. An important question will be whether the delayed subgroup has anything that distinguishes it from a normalbut delayed pattern. Measures of expressive language, rapid automatic naming, speed of processing printed letters and words, orthographic skill, and phonological coding in reading were administered as a means of further describing the profile of cognitive strengths and weaknesses in the phonological and delayed subgroups. The longitudinal nature of the design allowed us to investigate the stability of the subgroups over a one-year period and to begin to examine directly the deficit versus delay distinction.

METHOD

SUBJECTS

A total of 162 subjects were selected for this study from two cohorts totalling 230 children who were participating in a longitudinal study. Subjects who were not utilized in the present analyses generally had Wechsler Vocabulary scores that were too low, or Woodcock Word Identification scores that did not fit the range proscribed for a particular group. Data from assessments conducted in Fall 1996, 1997, and 1998 are reported in this paper. Normally achieving readers in grades 1, 2, and 3, and dyslexics in grade 3 were nominated by teachers at the outset of the study. After obtaining consent, screening criteria were applied. Children were not included in the study if they were rated by school staff as having limited English proficiency (these ratings were based on interviews with the child by a speech-language therapist). Other exclusionary criteria were severe cognitive or neurological impairments, severe hearing loss, or visual impairments, based on school records.

The dyslexic group consisted of 72 third-grade children with a mean age of 8 years, 6 months (range 7 years, 11 months to 9 years, 4 months). The group was more diverse than a typical dyslexic sample as it included some children with below average expressive language skills. The criteria for inclusion in the group was a scaled score on the WISC-III Vocabulary test (Wechsler 1992) of 7 or higher, and a score at or below the 26th percentile on the Woodcock Reading Mastery Test - Word Identification (Woodcock 1987).

A group of 44 same-age normal readers in the third grade (the CA group) was obtained as well. All were reading at or above the 40th percentile on the Woodcock Word Identification test and had a Vocabulary scaled score of 7 or higher. The mean age was 8;5 (range 7;11 to 9;3).

A total of 46 younger normal readers (the RL groups) were tested in January to March of the same academic years as the dyslexics and the CA group. There were two RL groups, one for comparison to the dyslexics in year 1 of the study, and one for comparison in year 2. The year 1 RL group consisted of 27 first and 6 second graders who had a very similar mean and range of Woodcock Word Identification scores as the dyslexic group. The mean age was 6;9 (range: 6;1 to 7;10). The RL group for year 2 consisted of 18 second graders from the year 1 RL group (who were first graders in year 1) and an additional group of 12 first graders. The mean age of this group was 7;3 (range: 6;1 to 7;11).

The initial identifying information for the subject groups in year 1 (mean and range for Woodcock Word Identification grade equivalent and percentile and WISC-III Vocabulary scores) is shown in table I. The normally achieving reader groups were higher in Vocabulary than the dyslexic group, a fact considered in the analyses described below.

Table I. Mean word identification and vocabulary scores at time 1 (standard deviations in parentheses).

	Woodcock Word Identification		WISC-III Vocabulary
Subject Groups	**Grade Equivalent**	**Percentile**	**Scaled Score**
Dyslexics (n = 72)	2.21 (0.36)	11.69 (7.85)	9.29 (1.80)
Reading Level (RL)			
Comparison Group (n = 33)	2.17 (0.39)	83.55 (13.92)	12.91 (2.82)
Chronological Age (CA)			
Comparison Group (n = 44)	4.07 (0.60)	70.02 (15.81)	10.54 (264)

PROCEDURES

Subjects were run individually over five 30-minute test sessions in 1996 (cohort 1) or 1997 (cohort 2) , and four 30-minute test sessions in 1997 or 1998. All testing was conducted at the schools during normal school hours.

Woodcock Reading Mastery Test - Word Identification. Form G was administered to children as a measure of sight word vocabulary in both years of the study. The items on the test are a representative sample of English words. About 25 percent of the words fit the general criteria for exception words utilized in the present study (for example, the pronunciation violates one or more common spelling-sound correspondences, or represents an orthographically "strange" combination of letters).

Pseudoword Reading. A list of 70 pseudowords was created for the study. The items ranged from simple CVC patterns (*nug*) to patterns with two or more letter clusters (*chome, scridge*). Some two-syllable items were included as well (*stining, namsion*). The items were ordered in difficulty based on pilot data. Children read the items aloud. Testing was discontinued when children made ten consecutive mistakes. The internal consistency reliability (Cronbach's alpha) was 0.96 for year 1. The task was administered in year 2 with a slight reordering of items.

Exception Word Reading. A list of 70 exception words also was created. The items were ordered in difficulty from easiest to hardest based on frequency and grade norms (Adams and Huggins 1986; Carroll, Davies, and Richman 1971) (e.g., *have, people, island, yacht, silhouette*). The task was discontinued when children made six consecutive mistakes. Internal consistency reliability for this task was high (Cronbach's alpha = 0.96).

Phoneme deletion. A task of the type devised by Bruce (1964) was administered in two parts. In part one, subjects repeated a familiar word that was spoken on a tape. The speaker on the tape asked the subject to repeat the word but with a specified part missing such as "snow" without the /s/ and "act" without the /k/. A single phoneme or a blend of two phonemes was deleted from the beginning, middle, or end of the word. There were 25 items, and testing was discontinued if the child made five mistakes in a row. Cronbach's alpha was 0.88. In part two, the items were all pseudowords such as "kimp" without the /m/. There were 15 items, and testing was discontinued if the child made five mistakes in a row. Cronbach's alpha for the pseudowords was 0.84. Scores were combined across the two tasks for the present analyses.

Wechsler Intelligence Scale for Children-III Vocabulary. The Vocabulary subtest of the WISC-III (Wechsler 1992) was administered. The task required children to define a series of words as best they could.

Clinical Evaluation of Language Fundamentals (CELF) Recalling Sentences. The Recalling Sentences subtest of the CELF (Semel, Wiig, and Secord 1995) was administered in year 2. Children listened to tape recorded sentences and repeated them back. Sentences gradually increased in length and grammatical complexity as well as word length and difficulty.

Three tasks were presented on laptop computers (Letter Matching, Orthographic Choice, and Semantic Categorization). For each task, the child viewed one or two letter strings or words on the screen and pushed a button to indicate the response. Responses were timed and latency was calculated for correct trials only.

Letter Matching. This task required children to decide if a string of five letters contained two letters that were the same or not and press a button (gtaVA = "yes"; pNhB = "no"). On some "same" trials, the matching letters were both in the same case (e.g., HntKt) and on others they were alternating case (e.g., gtaVA); therefore, the task required children to process letters to the point of recognition rather than making comparisons strictly on the basis of low-level physical features. Cronbach's alpha for the accuracy scores was .86. The procedure was based on a similar task employed by Bigsby (1988).

Orthographic Choice. This task required children to view two strings of letters displayed side by side on the screen and decide which item represented a correctly spelled word. They pressed a button to indicate which side of the screen had the correctly spelled item. Half of the items contained at least one exception word (*sponge, spunge*) and half contained two regular words (*sheap, sheep*). All of the foil items were identical phonologically to the correct exemplar items and hence the child could not rely on phonological recoding of the items to recognize the answer. This task taps the ability to recognize specific spellings of words. There were 48 items in all. Six practice trials were given. Cronbach's alpha for the accuracy scores was equal to 0.75, somewhat lower than the other tasks. The task was adapted from a similar measure used by Olson et al. (1989).

Semantic Categorization. For this task, children saw a printed category displayed on the screen (e.g., a fruit, a part of your body). The category was read aloud by the experimenter. It then disappeared from the screen and was replaced a second later by a word. The children decided if the item was a member of the category or not and pressed a button. There are four types of items that could be presented: actual exemplars (*pear*), alternative exemplars (*peach*), homonym foils (*pair*), and visual foils (*peer*). Only the last three types of items were shown. Subjects had to respond "yes" to alternative exemplars and "no" to foils. The task measures the extent to which subjects rely on phonological recoding when reading words for meaning. Children who are unable to reject homonym foils display evidence of relying on phonological recoding (accepting *pair* because it sounds exactly like *pear*). The visual foils were designed to share as many letters with the actual word (*pear*) as did the homonym foils to control for guessing based on visual similarity. There were sixteen categories and 40 items in all displayed (16 alternative exemplars, and 12 each of homonym foils and visual foils). There were 6 practice trials. Cronbach's alpha for the accuracy data was 0.58. The task was adapted from a similar measure developed by Van Orden for adults (Van Orden 1987) and by Sprenger-Charolles, Siegel, and Bechannec (1998) for children. The stimuli are listed in the Appendix.

Rapid Automatic Naming - Letters. Children named letters in an array as rapidly as they could and overall time was recorded. In year 1, the version of the RAN used by Torgesen et al. (1997) was presented. There were 36 letters arranged in four rows of 9. Only six unique letters were used (a, c, k, n, s, t), and these letters were ordered randomly and repeated in blocks of six. Two trials of 36 letters were presented. Test/retest reliability was 0.76. In year 2, the original Denckla and Rudel (1976) stimuli were used. There were five unique letters (a, d, s, o, and p) arranged in five rows of ten with random ordering within each block of five. Two trials of 50 letters were presented. Test/retest reliability was 0.78. Speeds in items per second were calculated to permit comparisons across years.

Title Recognition Test. Children were shown 45 book titles on a sheet and listened as the experimenter read them aloud one at a time. The child then rendered a judgement as to whether the book was a "real" book. They were encouraged to say "I don't know" if a book title was unfamiliar. There were 15 phony book titles. Children were warned that "some titles were not real books" and told not to guess. The total score on the task was the proportion of correct book titles chosen by the child minus the proportion of incorrect titles chosen by the child. This score corrects for guessing in the manner devised by Cunningham and Stanovich (1990). The task was designed to measure the general amount of print exposure for each child. The reliability for hits was .82 (Cronbach's alpha).

RESULTS

Analyses centered on three questions. First, how stable were the classifications of the dyslexics? Second, what cognitive deficits were associated with each subgroup pattern, particularly the delayed pattern? And third, how did the subgroup classification overlap with other leading subtyping schemes such as the phonological/surface dyslexia and double-deficit schemes?

Subgroup Definitions and Stability. Dyslexics were divided into subgroups in the first year of the study, based on the performance of the younger normal readers. For the Pseudoword Reading and Phoneme Deletion tasks, z-scores were created, based on the means and standard deviations for the RL group. Dyslexics were assigned to the phonological dyslexic subgroup ($n = 32$) if their score on either the Phoneme Deletion or Pseudoword Reading task fell at or below –0.9. Dyslexics were assigned to the delayed dyslexic subgroup ($n = 40$) if their score on both Phoneme Deletion and Pseudoword Reading was at or above –0.9. Since the younger normals were equated to the dyslexics on word recognition skill, this procedure results in one subgroup whose phonological skills are considerably lower than expected based on word recognition, and another whose phonological skills are on a par with their word recognition.

The distributions of Phoneme Deletion and Pseudoword Reading scores for year 1 are shown in figures 1 and 2 as a function of Woodcock Word Identification skill for the dyslexic subgroups, the RL group ($n = 37$), and the CA group ($n = 52$). Some data points overlap and are not shown in the graph. The graphs serve to make two important points that are basic to the present study. First, the relation between Word Identification and the phonological measures is strongly linear, consistent with the phonological core hypothesis. Second, there was considerable variability in phonological skill among the dyslexics, with some scoring as high as the CA-matched group. It is this variability that motivates an examination of subgroups of dyslexics. The distributions for the RL group and for the phonological group do not overlap completely, consistent with the definition of the subgroup, whereas the distributions for the delayed and RL group overlap very closely. Means on all tasks given in the first year are shown in table II.

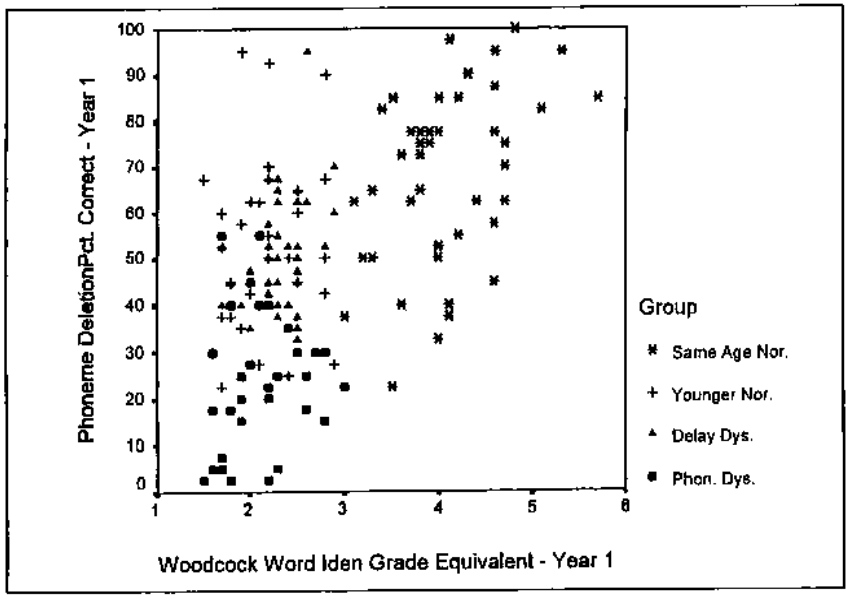

Figure 1. *Scatterplot of phoneme deletion scores as a function of Woodcock Word Identification grade equivalent scores in Year 1.*

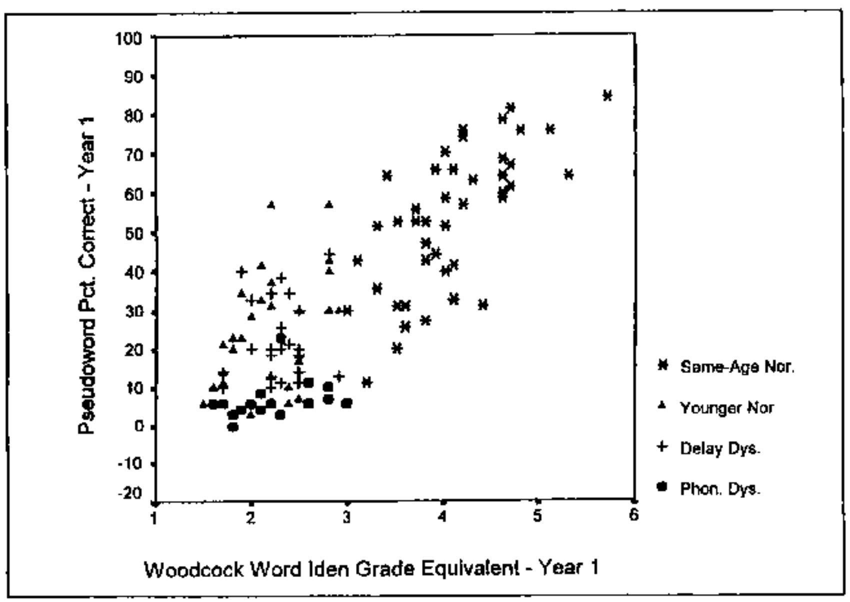

Figure 2. *Scatterplot of pseudoword reading scores as a function of Woodcock Word Identification grade equivalent scores in Year 1.*

Table II. Means, standard deviations and group differences for the test variables in year 1.

	Subject Groups			
	Dyslexic Subgroups		Comparison Groups	
Variables	Phonological (n = 32)	Delayed (n = 40)	RL Group (n = 33)	CA Group (n = 44)
Pseudoword Reading (%)	8.6 (6.8)	23.9 (11.0)	23.3 (14.6)	53.1 (17.8)
Phoneme Deletion (%)	23.4 (14.5)	50.8 (12.4)	52.5 (18.6)	67.2 (20.3)
Excep. Word Reading (%)	28.7 (12.4)	32.7 (9.2)	31.3 (12.5)	64.0 (10.4)
Orthographic Choice (%)	66.2 (10.3)	64.6 (9.5)	61.9 (13.2)	77.2 (8.0)
Letter Match (%)	83.6 (9.5)	86.7 (9.5)	82.8 (14.7)	82.6 (11.8)
Letter Match RT (in ms)	2434 (544)	2601 (696)	3055 (702)	2360 (671)
RAN-Letters (items/sec)	1.55 (0.29)	1.59 (0.27)	1.45 (0.24)	1.77 (0.32)

To determine whether the groups were distinct in year 1, a MANOVA was conducted on all of the measures from year 1, comparing phonological and delayed dyslexics and the RL group. The CA group was not included in this analysis as it was clearly superior to the other groups on the defining measures and the critical comparisons involved the groups matched on reading level. The groups differed significantly, $F(16, 170) = 8.32$, $p < .001$ (Pillai's trace, and at the .001 level on Wilks' Lambda and Hotelling's trace). Univariate ANOVAs revealed group differences on Letter Matching RT, Pseudowords, Phoneme Deletion, and Vocabulary scaled score. Post hoc tests (Tukey's test) indicated that the phonological subgroup performed more poorly than both the delayed subgroup and the RL group on Pseudoword Reading (p's < .001) and Phoneme Deletion (p's < .001). The phonological subgroup was lower on both measures. In addition, the phonological and delayed subgroups were lower than the RL group on Wechsler Vocabulary ($p < .05$), suggesting that the group differences on the reading tasks might be mediated by poor oral language ability. However, when Vocabulary was entered as a covariate in the MANOVA, this did not reduce the level of significance of group differences on the other measures. The three groups did not differ from each other on the exception, orthographic choice, and Woodcock tasks, indicating that their level of word-specific orthographic knowledge was comparable. These analyses established that the groups had clear operational definitions in year 1.

Since the Tukey test is conservative, it is possible that it might fail to detect differences that would be of interest. Therefore, t-tests also were conducted for each pairwise group difference among the dyslexics and RL group. The only additional group differences that were found were that both dyslexic subgroups were faster than the RL group on Letter Matching ($p < .01$) and the delayed group was faster than the RL group on RAN-Letters ($p < .05$).

The children's scores one year later on the same tasks, relative to the RL group for year 2 (n = 57), were used to classify them again. Of the 26 phonological dyslexics who returned for the second year of the study, 21 were in the same subgroup, which represents 80.8 percent stability. Similarly, of the 32 delayed dyslexics who returned, 27 were in the same subgroup for a stability of 84.4 percent. There were five cases in each group who switched subgroups over the one-year period. Most of these were children whose scores fell near the borderline for the subgroup division.

Correlations among the measures in year 2. Correlations among the measures given in year 2 are shown in table III for the dyslexic sample. It is important to point out that the current sample of dyslexics is highly selected and has a restricted range on many variables (particularly those relating to word recognition and phonological decoding) which may affect the correlations. Of particular interest was the pattern of relationships among the timed measures. Letter Match RT correlated with Orthographic choice RT, but neither of them correlated with RAN-Letters. Surprisingly, RAN-Letters did not correlate with Orthographic Choice accuracy, although it did with Woodcock Word Identification and Exception Word Reading. The phonological tasks (Phoneme Deletion and Pseudoword Reading) were correlated with each other, and Pseudoword Reading also correlated with word-level tasks such as Woodcock Word Identification, Exception Word Reading, and Semantic Categorization (alternative exemplars). These data indicate that phonological recoding is a very important part of reading words aloud, and silently for meaning among the dyslexic sample. Phoneme Deletion also correlated with Title Recognition. This surprising finding suggests that children with better phoneme awareness have higher print exposure. Title Recognition was

Table III. Correlations between the variables in year 2 (vocabulary is year 1) for dyslexics (n = 72). values over 0.27 are significant at $p < .05$.

Year 2 Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Woodcock Word Iden. Grade Equivalent														
2. Vocab. SS (yr 1)	0.17													
3. CELF Rec. Sent.	0.07	0.33												
4. Pseudoword Reading	0.52	0.14	0.13											
5. Phoneme Deletion	0.27	0.25	0.22	0.38										
6. Excep. Word Reading	0.81	0.10	0.02	0.37	0.01									
7. Ortho. Choice %	0.45	–0.35	–0.17	0.27	–0.10	0.47								
8. Ortho. Choice RT	–0.18	–0.05	–0.19	0.23	0.08	–0.26	–0.09							
9. Letter Match %	–0.12	–0.06	–0.17	0.03	0.10	0.00	0.11	0.24						
10. Letter Match RT	–0.06	–0.03	–0.04	0.25	–0.02	–0.14	0.00	0.40	–0.07					
11. RAN (items/sec.)	0.35	0.15	–0.02	0.11	0.11	0.35	0.05	–0.08	0.08	–0.19				
12. Title Recognition	0.27	0.22	–0.05	0.08	0.29	0.14	0.02	0.03	0.25	–0.26	–0.14			
13. Sem. Cat.- Alt. Ex.	0.56	–0.02	0.16	0.37	–0.03	0.62	0.50	0.04	0.14	0.14	0.11	–0.04		
14. Sem. Cat.- Hom. Foil	0.10	–0.05	–0.15	–0.20	–0.07	0.16	0.08	–0.10	0.03	–0.19	0.00	0.24	0.05	

not correlated with measures of word reading and orthographic skill, as has been the case in past studies with unselected samples of children (e.g., Cunningham and Stanovich 1990). CELF Recalling Sentences correlated with Vocabulary, but not with the phonological tasks, suggesting that it tapped into more general verbal skills rather than specific phonological skills.

Comparisons of the groups in year two. Analyses of the year 2 data were conducted on the groups as classified in year 1 (including the newly consituted RL group). The CA group had a 100 percent return rate, but five children had fallen below the 40th percentile on the Woodcock and were excluded from analyses.

The first question concerned whether the subgroups' test profiles were similar in years 1 and 2. Mean scores in year 2 on all of the tasks carried over from year 1 are listed in table IV.

Table IV. Means, standard deviations and group differences for the test variables in year 2.

	Subject Groups			
	Dyslexic Subgroups		Comparison Groups	
Variables	Phonological (n = 26)	Delayed (n = 32)	RL Group (n = 30)	CA Group (n = 39)
Pseudoword Reading (%)	22.3 (13.9)	35.4 (13.3)	39.7 (16.9)	63.4 (17.6)
Phoneme Deletion (%)	33.6 (14.9)	52.5 (13.6)	56.2 (19.2)	71.0 (17.7)
CELF Recalling Sent. SS	7.9 (3.2)	9.8 (2.8)	13.3 (8.9)	11.1 (4.8)
Excep. Word Reading (%)	50.8 (12.0)	49.2 (13.3)	45.4 (13.8)	74.4 (9.5)
Orthographic Choice (%)	72.2 (10.7)	71.9 (8.3)	72.0 (8.9)	82.1 (8.2)
Orthographic Choice RT	2045 (600)	2683 (1573)	2916 (1777)	1927 (472)
Letter Match (%)	85.9 (12.9)	88.2 (11.3)	92.0 (6.0)	91.3 (9.6)
Letter Match RT (in ms)	2139 (540)	2491 (622)	2680 (825)	2208 (629)
RAN-Letters (items/sec)	1.79 (0.29)	1.69 (0.27)	1.56 (0.22)	1.97 (0.30)
Title Recognition Score	.16 (.21)	.26 (.17)	.14 (.15)	.31 (.13)

Additional variables present in the second year were Orthographic Choice Latency, CELF Recalling Sentences scaled score, and Title Recognition Test corrected score. An initial MANOVA conducted on all of the variables is shown in table IV (the CA group was once again excluded from this analysis). There was a significant overall effect of group, $F(32, 86) = 3.189$, $p < .001$. Univariate ANOVAs were significant for CELF Recalling Sentences, Pseudoword Reading, Phoneme Deletion, and Letter Matching RT. Using Tukey post-hoc tests, phonological dyslexics differed from the RL group on the two phonological tasks and CELF Recalling Sentences (p values all less than .002), as well as from the delayed group (p values all less than .05). Thus, in addition to showing stability on the defining measures, phonological dyslexics performed poorly on an additional measure tapping linguistic processes. The fact that CELF Recalling Sentences correlated with Vocabulary but not Phoneme Awareness suggests that it taps into more general expressive language skills. Delayed dyslexics closely resembled the RL group on both phonological tasks, as well as the two measures of word-specific orthographic knowledge (Orthographic Choice and Exception words). The only other difference significant by Tukey post hoc test was that phonological dyslexics were faster than the RL group on the Letter Matching task ($p < .05$).

Planned contrasts using the t-test revealed several additional significant differences. Delayed dyslexics scored lower than the RL group on CELF Recalling Sentences ($p < .05$). Phonological dyslexics were faster on Letter Matching compared to both delayed dyslexics ($p < .05$) and the RL group ($p < .01$), and phonological dyslexics were marginally faster than delayed dyslexics on Orthographic Choice ($p < .06$) and marginally lower on the Title Recognition Test than the delayed dyslexics ($p < .07$). Delayed dyslexics scored higher than the RL group on the Title Recognition Test ($p < .05$). Phonological and delayed dyslexics did not differ on the RAN-Letters task.

The results indicate that the only salient area of difference favoring phonological dyslexics over delayed dyslexics is in the area of rate of processing, but this difference did not extend to the RAN task, limiting the generality of the deficit. There was no evidence that delayed dyslexics had a slower learning rate, based on the fact that they were not falling behind the phonological dyslexics on any of the reading or decoding tasks (e.g., Pseudoword and Exception Word Reading, Woodcock Word Identification, and Orthographic Choice), and indeed made as good or better progress than the phonological group on these tasks (see tables II and IV). In addition, delayed dyslexics did not differ from the RL group on any of these tasks.

Delayed dyslexics by definition had relatively good phonological skills, although they were still below the level of the CA group. It is of interest, therefore, to explore the extent to which they relied on phonology in reading for meaning. Previous studies have shown that reliance on phonology among normal readers increases with reading experience and with increases in decoding skills (e.g., Doctor and Coltheart 1980; Sprenger-Charolles et al. 1998). Delayed dyslexics may show appropriate levels of reliance on phonology for their reading level in keeping with their delayed profile. Alternatively, they may show enhanced reliance on phonological recoding, possibly implicating this as a compensatory mechanism. Reliance on a phonological recoding strategy in reading for meaning was assessed by means of the semantic categorization task. Readers who rely strongly on phonological codes in processing words for meaning are expected to find it difficult to reject a homonym for a word that fits the category prompted (e.g., flower–ROWS) (Sprenger-Charolles et al. 1998; Van Orden 1987).

Results for semantic categorization are shown in table V for all four groups. It is notable that all three groups had considerable difficulty rejecting the homonym foils in comparison to their performance on both the visual foils and the alternative exemplars. The visual foils provide a control for the possibility that dyslexics select homonyms because of a general tendency to process printed words only partially. An ANOVA comparing the phonological, delayed, and RL groups on all three conditions of the semantic decision task revealed significant main effects of group and condition as well as an interaction between group and condition, $F(4, 158) = 3.89$, $p < .01$. Planned contrasts

Table V. Mean percent correct in each condition of semantic categorization (standard deviations in parentheses).

	Subject Groups			
	Dyslexic Subgroups		Comparison Groups	
Variables	Phonological	Delayed	RL Group	CA Group
Alternative Exemplars	82.8 (12.3)	86.5 (11.5)	86.8 (8.9)	92.8 (7.4)
Visual Foils	75.7 (19.3)	81.6 (17.5)	87.1 (13.2)	93.8 (8.2)
Homonym Foils	51.8 (14.9)	43.9 (15.1)	54.2 (20.1)	59.5 (18.2)

by t-test indicated that the phonological dyslexics had a lower score than the RL group on the visual foils ($p < .025$) and the delayed dyslexics had a lower score than the RL group on the homonym foils, ($p < .05$). Delayed dyslexics showed a trend toward lower scores on the homonym foils than the phonological dyslexics ($p < .06$). The results indicate that phonological dyslexics make more guesses based on visual approximations than the RL group, but that delayed dyslexics showed a stronger tendency to rely on phonological codes when reading for meaning.

Overlap between the current subtyping scheme and other subtyping schemes. The analyses up to this point indicate that, aside from a greater reliance on phonological coding in reading for meaning (see Discussion for further comments), the delayed dyslexics are very similar to the RL group in both years of the study. However, a question that arises is whether there are salient subgroups *within* the general category of delayed dyslexics that have more distinctive cognitive profiles. Two subtyping schemes are relevant to the measures utilized in this study: the phonological/surface distinction and the double-delay scheme. Both schemes are explored below.

Earlier in the paper, it was pointed out that surface dyslexics, who have poor Exception Word Reading and relatively good phonological skills, would fall, by definition, into the delayed subgroup. We identified nine surface dyslexics in year 1 who met the criteria for the delayed subgroup, and in addition, showed an Exception Word Reading z-score below zero and a discrepancy of at least .5 standard deviations between Exception Word and Pseudoword Reading. Mean scores on the variables for both years are shown in tables VI and VII.

Six of the eight subjects returning in year 2 were classified as delayed and two were classified as phonological dyslexics. Only two of the surface dyslexics were reclassified as surface dyslexics in year 2. Therefore, this subgroup is very unstable. It is apparent from tables VI and VII that despite their low scores on Exception Word Reading and relatively high scores on Pseudoword Reading in year 1, the surface dyslexics were largely indistinguishable from the RL group in year 2. T-tests comparing surface dyslexics in year 2 with phonological dyslexics revealed differences in Pseudoword Reading and Phoneme Deletion favoring the surface dyslexics (p values less than .025) , but no differences in Exception Word Reading, Letter Matching, Orthographic Choice, RAN-letters, and Semantic Categorization. Surface dyslexics were lower than

Table VI. Means, standard deviations and group differences, including surface dyslexics, for the test variables in year 1.

	Subject Groups			
	Dyslexic Subgroups		Comparison Groups	
Variables	Phonological ($n = 32$)	Surface ($n = 9$)	RL Group ($n = 33$)	CA Group ($n = 44$)
Pseudoword Reading (%)	8.6 (6.8)	29.5 (8.7)	23.3 (14.6)	53.1 (17.8)
Phoneme Deletion (%)	23.4 (14.5)	42.8 (5.2)	52.5 (18.6)	67.2 (20.3)
Excep. Word Reading (%)	28.7 (12.4)	21.4 (3.8)	31.3 (12.5)	64.0 (10.4)
Orthographic Choice (%)	66.2 (10.3)	59.4 (12.4)	61.9 (13.2)	53.1 (17.8)
Letter Match (%)	83.6 (9.5)	85.4 (8.8)	82.8 (14.7)	82.6 (11.8)
Letter Match RT (in ms)	2434 (544)	2700 (552)	3055 (702)	2360 (671)
RAN-Letters (items/sec)	1.55 (0.29)	1.59 (0.23)	1.45 (0.24)	1.77 (0.32)

the RL group on the homonym foil trials of the Semantic Categorization task ($p < .05$), but did not differ reliably from the RL group on any of the other measures. The most parsimonious interpretation of the data is that the surface group resembles the delayed dyslexic category.

The correspondence between the present subtyping scheme and the double-deficit scheme was explored by classifying subjects into four groups in year 1: (1) low phoneme deletion; (2) slow RAN time; (3) a double deficit; or (4) a deficit on neither task. A cutoff of one standard deviation below the CA group's mean was utilized. Table VIII reveals that most of the phonological dyslexic subgroup fell into either the phonological deficit or double-deficit group. In contrast, delayed subjects either fell into the slow RAN group or the no deficit group, with the majority falling in the no deficit group. By way of comparison, most of the CA subjects were in the no deficit group, although 7 were in the slow RAN group. These data indicate that roughly half of the slow RAN subjects were normal readers on an accuracy measure (the Woodcock Word Identification test), a result which parallels Morris et al.'s (1998) findings.

Table VII. Means, standard deviations and group differences, including surface dyslexics, for the test variables in year 2.

	Subject Groups			
	Dyslexic Subgroups		Comparison Groups	
Variables	Phonological ($n = 26$)	Surface ($n = 9$)	RL Group ($n = 30$)	CA Group ($n = 39$)
Pseudoword Reading (%)	22.3 (13.9)	35.0 (15.5)	39.7 (16.9)	63.4 (17.6)
Phoneme Deletion (%)	33.6 (14.9)	55.0 (21.6)	56.2 (19.2)	71.0 (17.7)
CELF Recalling Sent. SS	7.9 (3.2)	10.8 (1.8)	13.3 (8.9)	11.1 (4.8)
Excep. Word Reading (%)	50.8 (12.0)	42.1 (15.4)	45.4 (13.8)	74.4 (9.5)
Orthographic Choice (%)	72.2 (10.7)	68.0 (9.4)	72.0 (8.9)	82.1 (8.2)
Orthographic Choice RT	2045 (600)	2175 (864)	2916 (1777)	1927 (472)
Letter Match (%)	85.9 (12.9)	85.9 (8.7)	92.0 (6.0)	91.3 (9.6)
Letter Match RT (in ms)	2139 (540)	2231 (362)	2680 (825)	2208 (629)
RAN-Letters (items/sec)	1.79 (0.29)	1.71 (.32)	1.56 (0.22)	1.97 (0.30)
Title Recognition Score	.16 (.21)	.21 (.14)	.14 (.15)	.31 (.13)

Table VIII. Correspondence between the phonological and delayed classification and the double deficit classification in year 1.

	Double Deficit Classification			
Phonological/Delayed Groups	Low Phoneme Awareness	Slow RAN	Double Deficit	No Deficit
Phonological Dyslexics	15	2	10	5
Delayed Dyslexics	0	8	1	31
CA Group	3	7	0	34

DISCUSSION

Previous empirical studies, and the results of connectionist simulations of developmental dyslexia (Harm and Seidenberg 1999), suggest that dyslexic children are heterogeneous but that they might be parsimoniously placed into two general categories: phonological dyslexia, and delayed dyslexia. The goal of the present study was to present a rationale for this subgrouping scheme and to report initial data from an ongoing longitudinal investigation. The study yielded results that shed light on three empirical issues: (1) the stability of the subgroups; (2) the nature of correlated reading and language deficits in the subgroups; and (3) the developmental trend over one year.

To summarize, phonological and delayed dyslexics both showed better than 80 percent stability over one year. They were distinguishable from each other, as well as from younger normal readers, on several tasks. Phonological dyslexics continued to be poor one year later on the classification tasks, and on an additional expressive language task, CELF Recalling Sentences. Delayed dyslexics showed evidence of slow processing of letter-level and orthographic information. Their latencies were comparable to the RL group, who were one and a half years younger on average. Delayed dyslexics also showed evidence of relying more strongly on phonological codes in the Semantic Categorization task. Despite these differences, the overall profile across the test battery resembled that of the RL group, supporting the designation of this subgroup as delayed but normal. The amount of progress on Woodcock Word Identification over one year was roughly comparable for the two subgroups, indicating that delayed dyslexics were not simply learning to read at a slower rate than the phonological dyslexics. If anything, phonological dyslexics made slightly less progress in overall word reading over the year.

The findings on stability are the first to be reported for these particular subgroups of reading difficulty. Longitudinal data reported by Torgesen et al. (1997) are consistent with the present data as they indicated that phonological deficits (e.g., phonological awareness and pseudoword reading) were quite stable in a large sample of dyslexics, although the children in that study were selected for low phonological decoding ability. Studies by Snowling, Goulandris, and Defty (1996) and Manis, Custodio, and Szeszulski (1993) found that phonological difficulties tended to persist and even to become more distinct from word reading skill over a two-year period in unclassified samples of dyslexic children.

A central issue concerns whether the classification scheme presented here provides a clearer understanding of the variability among dyslexic children than other potential subtyping schemes. It is important to remember that any dyslexic classification scheme makes arbitrary distinctions in what is largely a continuous distribution of scores (Castles et al. 1999). Subgrouping is a tool to understand the variability among dyslexics, not an end in itself, and subtyping schemes can change as theories of the cognitive deficits underlying dyslexia evolve. The present scheme was derived from an explicit computational model of developmental dyslexia that produces profiles in simulation studies that resemble the present two subgroups.

We explicitly compared our classification scheme to two of the current approaches, the phonological/surface distinction (e.g., Castles and Coltheart 1993; Manis et al. 1996; Stanovich et al. 1997) and the double-deficit framework (Bowes and Wolf 1993; Wolf and Bowers in press). Surface dyslexics were unstable over the one year of the study, and were found to be a subset of the delayed subgroup who did not show any characteristics that would distinguish them from the larger group. Thus, while it is possible that surface dyslexia has unique distinguishing features, this was not apparent from the tasks given in the present study. The children were also crossclassified as having single or double deficits using Bowers and Wolf's double-deficit framework. The two frameworks overlapped to some extent as many phonological dyslexics were placed in either the double-deficit or single phonological-deficit category. The delayed dyslexics were mostly classified as having no deficit in the Bowers-Wolf scheme, but a few of them had single deficits in naming speed. Half of the children in the slow naming speed group were normal readers. The two approaches thus appear to be applicable to somewhat different populations of poor readers, although there is significant overlap.

This brings us to the most puzzling issue in the study. How do we characterize the delayed dyslexic subgroup? Recall that in the connectionist model, this group may be a hodgepodge of different deficits, ranging from cognitive deficits (visual-orthographic deficits) to experiential problems (low print exposure), but no one profile may dominate enough to produce a distinctive pattern of test performance. It is possible that such a state of affairs existed in the present sample, but this will require in-depth analysis of individual cases. What can be said about the delayed subgroup at present is that they were slow on two visual-orthographic tasks relative to phonological dyslexics, and that they relied to an unusual extent on phonological recoding in the semantic categorization task. While further testing of delayed dyslexics is necessary to determine whether these patterns are indeed robust, some preliminary hypotheses about the cognitive profile we observed can be evaluated.

First, are delayed dyslexics actually phonological dyslexics who have been partially or largely remediated? They fit the overall profile reported in past studies for remediated phonological dyslexics (e.g., Olson, Wise, Johnson, and Ring 1996; Torgesen, Wagner, and Rashotte 1997): relatively good phonological awareness and decoding in the context of continued slow and inaccurate word recognition. However, the intervention histories we obtained from the children's teachers indicate that if anything, the phonological dyslexics received more remediation. Sixteen of the 32 phonological dyslexics received such intervention during the third grade as did 13 of the 40 delayed dyslexics. The intervention most commonly involved 50 minutes, four days a week of small group, phonics-oriented instruction, although some children received less. Therefore, the data do not support the hypothesis that the delayed group was a remediated form of the phonological group.

Second, delayed dyslexics might represent a group of children with mild phonological deficits and low print exposure as proposed by Stanovich et al. (1997) for their sample of surface dyslexics. Delayed dyslexics in our sample clearly did not have low print exposure as measured by the Title Recognition Test, but this test may be somewhat unreliable (Castles et al. in press). Delayed dyslexics did appear to have mild phonological deficits as they were low relative to the CA group but not the RL group. However, delayed dyslexics had about the same degree of impairment as phonological dyslexics on various word reading tasks such as the Woodcock Word Identification, Exception Word Reading, and Orthographic Choice. Therefore, mild phonological deficits *alone* are not sufficient to explain the delayed group's deficit. In addition, because reading and phonological skill are most likely reciprocally related, phonological deficits could be a by-product of slow progress in overall reading.

A third possibility is that delayed dyslexics have inefficient word decoding and recognition as indicated by their slow performance on some tasks, and by the observation that they have learned to rely more strongly on phonological recoding (e.g., in the semantic categorization task). The possibility that at least some delayed dyslexics have not developed rapid and

automatic word recognition skills is being explored in the third year of the study.

Finally, it is possible that delayed dyslexics represent a group of poor readers who lag behind good readers (for any of a number of reasons, such as poor preparation for schooling), but who will eventually catch up. Although there was no evidence of catch-up in the present study, the one-year period of study does not provide enough of a time span to really test this hypothesis.

CONCLUSION

There is a great deal of interest in the nature and sources of variability within the dyslexic population. The present study utilized a connectionist model of word reading (Harm and Seidenberg 1999) to explore the variability question. We defined two subgroups of dyslexics on the basis of accuracy-based measures of reading. Phonological dyslexics had deficits in nonword reading and phoneme awareness, in relation to an RL group equated on word recognition, whereas delayed dyslexics had phonological skills on a par with the RL group. The subgroups were found to be moderately stable (82 percent) and to be distinct on a number of theoretically interesting measures. The subgroups showed a different profile of deficits on measures of oral language, orthographic skill, and reliance on phonology in reading for meaning. Phonological dyslexics had isolated deficits in phonology and expressive language, whereas delayed dyslexics resembled the RL group across all reading tasks. The delayed group was not synonymous with general cognitive delays or general deficits in rate of processing. Instead, the data suggest that this subgroup of dyslexics has difficulties with the development of orthographic skills, including the acquisition of rapid, automatic word recognition, and possibly differences in word recognition strategies. The data validate the phonological/delayed distinction proposed by Harm and Seidenberg (1999), and highlight the need to explore the nature and source of the difficulty in delayed dyslexics.