Varieties of Developmental Reading Disorder: Genetic and Environmental Influences

Anne Castles

University of Melbourne, Parkville, Victoria, Australia

and

Helen Datta, Javier Gayan, and Richard K. Olson

University of Colorado at Boulder

There is widespread support for the notion that subgroups of dyslexics can be identified who differ in their reading profiles: Developmental phonological dyslexia is characterized by poor nonword reading, while developmental surface dyslexia is distinguished by a particular difficulty in reading irregular words. However, there is much less agreement about how these subtypes, and particularly the surface dyslexic pattern, are to be accounted for within theoretical models of the reading system. To assist in addressing this issue, the heritability of reading deficits in dyslexic subtypes was examined using a twin sample. Subjects' scores on (a) an exception word reading task and (b) a nonword reading task were used to create a subtype dimension, and surface and phonological dyslexic subgroups were selected from the ends of this distribution. Reading deficits were found to be significantly heritable in both subgroups. However, the genetic contribution to the group reading deficit was much greater in the phonological dyslexics than in the surface dyslexics. The finding of differential genetic etiology across subtypes suggests that there is at least partial independence in the development of the cognitive processes involved in reading exception words and nonwords. Also, the results support accounts of surface dyslexia which emphasize a strong environmental contribution. © 1999 Academic Press

Key Words: reading disorders; dyslexia; subtypes of dyslexia; genetic influences; behavioral genetics; orthographic deficits; phonological deficits.

Many reading researchers have questioned the notion that developmental reading disorders occur in only one form. A complex and multifaceted process such as reading, it would seem, will be likely to fail in an equally complex and multifaceted range of ways. Consequently, there has been a relatively long history of attempts to classify reading disorders into different categories or "subtypes" (e.g., Boder, 1973; Fletcher & Morris, 1986). However, as Stanovich, Siegel, and Gottardo (1997) have recently noted, much of the earlier work in subtyping was somewhat disappointing, since it was not based on explicit models of the reading system and therefore tended to be descriptive rather than explanatory. More recently, researchers have made predictions about patterns of developmental reading disorder based on precise and, in some cases, fully implemented models of the skilled reading system (Castles & Coltheart, 1993; Coltheart, Curtis, Atkins, & Haller, 1993; Manis, Seidenberg, Doi, McBride-Chang, & Peterson, 1996; Olson, Kliegl, Davidson, & Foltz, 1985; Plaut, McClelland, Seidenberg, & Patterson, 1996). This has led to renewed interest in the basis of the subtypes themselves and in their potential for adjudicating between competing models of the reading process. The aim of this paper is to report some recent work which bears on both of these issues. Specifically, we discuss some results seeking to validate the subtypes identified by Castles and Coltheart (1993) and to assess genetic influences on reading deficits in these subtypes.

The large-scale study of varieties of developmental dyslexia conducted by Castles and Coltheart (1993) was conceived of as a means of providing support for dual-route models of the skilled reading system (e.g., Coltheart, 1978; Coltheart et al., 1993; Morton & Patterson, 1980). These models propose that skilled readers have at their disposal two distinct procedures for reading aloud: A lexical or word-specific procedure, which involves gaining access to internal units representing whole words, and a nonlexical procedure, which involves using a system of rules that specify relationships between subword units, such as graphemes and phonemes. Castles and Coltheart reasoned that, if these procedures indeed function independently in the skilled reading system, they should be capable of being separately developed in children learning to read. Therefore, two patterns of reading disorder should be identifiable: Developmental phonological dyslexia, where there has been a selective deficit in developing the nonlexical procedure, and developmental surface dyslexia, where there has been a selective deficit in acquiring the lexical procedure. The former would be able to be identified through specifically poor reading of nonwords (e.g., *giph*), since reading such items is presumed to require the use of conversion rules, while the latter would be identified through specifically poor reading of exception words (e.g., *yacht*), since success with these items is thought to require access to word-specific information.

Castles and Coltheart (1993) tested a sample of 53 poor readers on their ability to read aloud sets of irregular words and nonwords. Based on their scores on these tasks, eight subjects were identified as pure developmental phonological dyslexics: Their nonword reading was poor, compared with chronological age-

matched controls, but their exception word reading was within normal range. Another 10 subjects were classified as pure developmental surface dyslexics: Their exception word reading was poor but their nonword reading fell within normal range. A further 27 subjects were poor on both tasks and were therefore not classified as "pure" cases, but nevertheless showed a significant discrepancy between their scores on the exception word and nonword tasks. Castles and Coltheart concluded that these results were best interpreted in terms of a dual route model, with the subtype profiles representing different levels of development of the lexical and nonlexical procedures.

Recently, Manis and colleagues (Manis et al., 1996) have extended this work in several important ways. Firstly, they provided some validation of the subtypes by showing that subjects in the two subgroups differed in predicted ways on tasks that would be expected to be related to their respective reading deficits but which were independent of the actual tests used to classify them (i.e., exception word and nonword reading). Specifically, they found that the phonological dyslexics had more difficulty analyzing the phonemic structure of spoken nonwords than either the surface dyslexics or a reading level-matched control group, while they were better at making discriminations based on orthography than either of the comparison groups. In addition, the phonological dyslexics were less likely than the comparison groups to produce erroneous responses that were phonologically appropriate, but were more likely to produce visual approximations to target words. Such findings support the conclusion that there are indeed distinct patterns of developmental dyslexia at the extreme ends of the distribution of performance.

Manis et al.'s second major finding was that, although the phonological dyslexic subjects were poorer at nonword reading than both chronological and reading age-matched controls, the surface dyslexic subjects were no poorer at exception word reading than the reading level-matched controls. They therefore concluded that the surface dyslexic pattern is best described as a general developmental delay rather than as a "deviant" reading pattern. Based on the connectionist models of Seidenberg and McClelland (1989) and Plaut et al. (1996), they argued that this delay might be attributable to a basic "resource limitation," which results in slower mastery of all reading stimuli. This limitation may be modulated by other factors, such as degree of text exposure or emphasis on phonics instruction in the curriculum.

Stanovich, Siegel, and Gottardo (1997) report a similar finding to Manis et al. (1996) in relation to surface dyslexics, showing that these subjects display a cognitive profile on a range of tasks that is very similar to that of reading level-matched controls. Taking a different approach to that of Manis et al., they suggest that this delay might be due, not to a resource limitation, but to mildly depressed phonological skills combined with exceptionally inadequate reading experiences. They note that a higher proportion of surface dyslexics was found in their sample, which was drawn from low achieving schools, than was found in the sample of Castles and Coltheart (1993). Insofar as a low-achieving school

might serve as a proxy for inadequate text exposure, they argue, this result can be seen as consistent with the hypothesis of a relationship between surface dyslexia and inadequate print exposure.

Early reading instruction is a third possible influence on the expression of subtypes in dyslexia. Phonological training programs for children with dyslexia have been shown to have strong beneficial effects on both phoneme awareness and phonological decoding (nonword reading), although gains in rapid word recognition are more modest (Olson, Wise, Ring, & Johnson, 1997). Most phonologically trained children in this study had a phonological dyslexic profile at the start of training when compared to younger normal children at the same reading level (Rack, Snowling, & Olson, 1992). Most of these children shifted to a surface dyslexic profile by the end of training and in follow-up testing 1 year after the end of training. However, this study only shows that dyslexics' positions on a subtype dimension can be changed through intense phonological remediation. It does not necessarily explain the etiology of subtype differences or reading deficits among dyslexics selected randomly from the population, where genetic and environmental factors may be involved.

In summary, it would seem that there is now general agreement that it is possible to identify subgroups of dyslexics who differ in their reading profiles. All of the studies above report evidence for surface and phonological dyslexic subgroups. However, there is much less agreement about how these profiles, and particularly the surface dyslexic pattern, are to be accounted for within theoretical models of the reading system. Some argue for a specific deficit in the use of the lexical procedure within a dual-route model (Castles & Coltheart, 1993), others for a general resource limitation within a connectionist network (Manis et al., 1996), and still others for a phonological deficit that is exacerbated by lack of exposure to reading materials (Stanovich et al., 1997).

An examination of genetic influences on reading deficits in subtypes of developmental dyslexia may assist in addressing these issues by providing converging evidence on the distinctiveness and etiology of the different reading patterns. Let us consider first the case of developmental surface dyslexia. An account of this subtype based on a general resource limitation, wherein reduced computational resources are modeled in terms of a reduction in hidden units in a connectionist network (Manis et al., 1996), would be compatible with both genetic and environmental influence (i.e., anoxia at birth, lead exposure, severe early deprivation). On the other hand, an account of surface dyslexia that implicates a lack of reading experience would seem to be more consistent with the prediction of a strong environmental component to the reading problem: Text exposure is presumably a factor which affects a child's experiences once he or she begins the task of learning to read.

The predictions about genetic influence on surface dyslexia that would follow from a dual-route model are difficult to specify: The dual-route cascading model of Coltheart et al. (1993) does not learn, so it is not yet obvious what processes

might be involved in the development of the lexical procedure and whether these are likely to be more genetically or environmentally based. However, in more general terms, the finding that surface and phonological dyslexia have very different profiles of genetic influence may provide some support for a dual-route model, in that it would suggest that there is substantial independence in the cognitive processes involved in reading exception words and nonwords.

In the case of phonological dyslexia, the different models would appear to converge on similar predictions of genetic influence. There is now widespread agreement that poor nonword reading is associated with underlying deficits in perceiving and manipulating the sounds of spoken language, or *phonological awareness* deficits (Bowey, Cain, & Ryan, 1992; Bradley & Bryant, 1978; Bruck, 1992; Liberman & Shankweiler, 1985; Manis, Custudio, & Szezulski, 1993; Pennington, Van Orden, Smith, Green, & Haith, 1990). Such deficits appear to some extent to be identifiable prior to learning to read (Wagner, Torgesen, & Rashotte, 1994) and may also be associated with subtle abnormalities in cortical structure (Galaburda & Livingstone, 1993). It would therefore seem likely that they have a significant genetic component, a hypothesis that has been supported by analyses of the heritability of phonological deficits (Olson, Wise, Conners, Rack, & Fulker, 1989; Olson, Forsberg, & Wise, 1994a). This genetic influence would be expected regardless of whether poor nonword reading is described in terms of an impaired nonlexical route (Castles & Coltheart, 1993) or in terms of degraded phonological representations in a connectionist network (Manis et al., 1996). Nevertheless, although it may not discriminate between competing models, the finding that phonological dyslexics show a strong heritability of their reading deficit would provide further evidence for an account of the disorder based on a core phonological processing deficit.

Olson and colleagues have been addressing issues of genetic influence on reading subskills for several years using twin data from the Colorado Learning Disabilities Research Center (Olson et al., 1989, 1994a). The logic underlying the twin study design is as follows: Monozygotic (MZ) twins share 100% of their genes, while dizygotic (DZ) twins share only 50% of their segregating genes, on average. MZ and DZ pairs have a similar shared environment in the home and school. Therefore, the degree to which MZ twins are more similar to each other than DZ twins can provide an estimate of genetic influence. Using a twin study methodology based on this logic, Olson and colleagues have found evidence for a significant genetic component to the group deficit in phonological skills (Olson et al., 1989) and, more recently, also to the group deficit in orthographic skills (Olson et al., 1994a).

The evidence for a genetic component to phonological and orthographic reading skills has come largely from studies of poor readers who have not been differentiated by subtype. Recently, Olson and colleagues have also begun to examine the relative heritability of reading deficits in developmental dyslexic subtypes using orthographic and phonological measures from the Colorado twin project (Gayan, Forsberg, & Olson, 1994; Gayan, Datta, Castles, & Olson, 1997). This work has provided some

preliminary support for the proposal that phonological dyslexics may show a stronger heritability of reading deficits than surface dyslexics.

The aim of the present study was to extend this work by examining the heritability of reading deficits in subtypes of developmental dyslexics chosen on the basis of measures similar to those used by Castles and Coltheart (1993). Subgroups of dyslexics were selected from the Colorado twin sample who were either (a) better at exception word reading than nonword reading (the phonological dyslexic group) or (b) better at nonword reading than exception word reading (the surface dyslexic group). Following from the work of Manis et al. (1996) and Stanovich et al. (1997), the external validity of these group selections was assessed by examining the performance of the two groups on other reading and language measures. Analyses of the heritability of the reading deficits in the two subgroups were then carried out.

METHOD

Subjects

A subset of data from the twin sample of the Colorado Learning Disabilities Research Center (CLDRC) was analyzed for this study. This sample of twins and their families has been accumulated since 1982 from 27 school districts across the state of Colorado. First, school records are used to identify all twin pairs. Then, those twin pairs in which at least one of the twins has a history of reading problems are invited to the laboratory to undergo an extensive battery of psychometric tests. A comparison group of twin pairs in which no member of the twin pair has a history of reading problems are selected as controls and are assessed on the same battery of tests.

The reading-disabled (RD) sample selected for the present study consisted of 967 children, who were chosen from the larger twin sample on the basis that they had a score of more than one standard deviation below age average (based on the control group) on a composite measure of word recognition. The composite was derived in equal weights from the subjects' scores on two single word reading tests: The Time-Limited Word Recognition Test (TLWRT) of Olson and colleagues (Olson et al., 1989, 1994b), which measures reading accuracy for words presented on a computer screen where correct responses are initiated within 2 s, and the Word Recognition subtest of the Peabody Individual Achievement Test (PIAT). This composite measure has proved to be a more reliable index of word recognition than either measure alone (Olson et al., 1994b). The score was thus designed to identify children whose basic word recognition skills, at the single word level, were deficient for their age. In some cases both members of a twin pair fell into this RD group, while in other cases only one member was included. In total there were 592 pairs of twins (272 MZ; 320 DZ) in which at least one twin of the pair fell into the RD category (this figure produces a larger number of subjects overall than the 967 listed above because some twins were entered

twice into the analysis). Of the DZ twins, 187 pairs were same sex and 133 pairs were opposite sex.

The children included in the RD sample ranged in age from 8 to 18 years ($M =$ 12 years, 2 months). Their performance was compared with that of a control group of normal readers, for the purpose of standardizing the scores relative to age. The control group consisted of 846 children, none of whom had a history of reading problems (423 twin pairs: 211 MZ; 212 DZ) and who ranged in age from 8 to 18 years ($M =$ 12 years, 2 months). Exclusionary criteria prescribed that the subjects in both the control and the RD groups have a verbal or performance IQ on the Wechsler Intelligence Scale of above 85, that they have no obvious neurological or sensory deficits, and that English be the first language spoken in the home.

Measures

In order to select phonological and surface dyslexic subgroups from the RD sample, the subjects were assessed on measures of phonological or "nonlexical" processing and orthographic or "lexical" processing. Their scores on these measures were subsequently used to create a subtype variable, which represented the discrepancy between their performance on the two tasks. RD subgroups were selected on the basis of this subtype variable. As well as this, the subjects' performance on a range of associated reading and language tasks was examined for the purpose of validating the selection of the subgroups. These measures included an orthographic choice task, a phonological choice task, a phoneme deletion task, and measures of error type in word reading. Details of all of these measures are provided below.

Subtyping Measures

Phonological processing. As was the case in Castles and Coltheart (1993), the phonological processing measure adopted was a nonword oral reading task. Part of the standard battery of tests undertaken by children visiting the Colorado laboratory, this task requires subjects to read aloud a set of 85 nonwords of varying difficulty (e.g., *per, strale, lobsel*) from lists developed by Olson et al. (1989, 1994b). For the present study, each subject's percent correct score on these items was used and this figure was converted into a standardized *z* score for the subject's age based on the scores of the control group.

Orthographic processing. Following Castles and Coltheart (1993), an ortho-graphic processing measure based on the subjects' accuracy in reading aloud irregular words was used. Specifically, 40 very irregular or "strange" words (Seidenberg, Waters, Barnes, & Tanenhaus, 1984) were identified from the Time Limited Word Recognition Test (TLWRT; Olson et al., 1989, 1994b) and a score for each subject was derived based on their accuracy in reading aloud these items. Strange words, as defined by Seidenberg et al. (1984), are those exception words which have very unusual orthographic structures, such as *yacht, fruit,* or *eye.* This was operationalized in the present study by using a combination of regularity and

bigram frequency (Solso & Juel, 1980): The selected items were all irregular and had a mean summed positional bigram frequency of 3417, which was significantly lower than the mean of 6125 for the other items in the TLWRT, t (194) = 3.14, $p < .01$. Our expectation was that, as such items require unique word-specific knowledge to be pronounced correctly, accuracy on these items was likely to be a good index of orthographic reading processes. The list of strange items can be found in the Appendix.

Subjects generally do not read all the words in the TLWRT: Basal levels are set using screening items and the test is discontinued once subjects have failed to initiate a correct response within 2 s on 10 of the last 20 items (or when the end of the list is reached). Thus, not all subjects read the same strange items, nor did they read the same number of items overall. For this reason, the orthographic reading score obtained for each subject was the number of strange items that the subject read correctly as a percentage of the total number of strange items that he or she was presented with. Once obtained, this percentage was again converted into an age-regressed standardized z score based on the performance of the control subjects. Overall, subjects read a mean of 11 strange items ($SD = 3.3$).

Validation Measures

Phonological choice. As a supplementary measure of phonological reading skills, the subjects' performance on the phonological choice task of Olson et al. (1994b) was assessed. In this task, subjects are presented with a group of three nonsense words (e.g., *kep, kap, ket*) and are asked to nominate which item sounds like a word. Skill in grapheme to phoneme conversion is assumed to be required for success in this task, since the correct answer cannot be obtained based on orthographic information. Therefore, some validation of the group allocations would be provided if the surface dyslexic group were found to perform better on the task than the phonological dyslexic group. Subjects' accuracy scores on this test were expressed as standardized z scores, relative to the performance of the control group.

Orthographic choice. The orthographic choice task of Olson et al. (1985, 1994b) was used as a further measure of orthographic processing. In this task, subjects are presented with a word (e.g., *rain, salmon*) and a pseudohomophone of that word (e.g., *rane, sammon*) and are asked to nominate which of the two is a word. Since use of grapheme to phoneme conversion rules will not allow a subject to differentiate between the two alternatives, use of word-specific information is assumed to be necessary to obtain the correct answer in this test. In this case, therefore, phonological dyslexics would be expected to perform better on the task than surface dyslexics. Subjects were presented with 80 pairs of items and pressed a left or right button to designate their response as quickly as possible after a pair appeared on the computer screen. Feedback was given for latency and accuracy after each trial. Once again, subjects' accuracy scores were converted into age-regressed standardized z scores based on the results of the control group.

Phoneme deletion. As discussed, there is mounting evidence that deficits in grapheme to phoneme conversion are associated with underlying problems in per-

ceiving and manipulating the sounds of spoken words (Bruck, 1992; Liberman & Shankweiler, 1985; Rack, Snowling, & Olson, 1992). Therefore, as a further validation measure, subjects were assessed on their performance on a spoken language phonological awareness task. The expectation here was that the phonological dyslexics would have more difficulty on the task than the surface dyslexics.

Specifically, a phoneme deletion task was used (Olson et al., 1994a), which was based on the tests of Bruce (1964) and Rosner and Simon (1971). The task consisted of 6 practice and 40 test trials. First, subjects heard a nonword, which they were asked to repeat. Then they were asked to remove a specified phoneme from the nonword which, if done correctly, resulted in a word (e.g., "say prot," "now say prot without the /r/ sound"—"pot"). The task was presented via a tape player. Two seconds were allowed for the subjects to repeat the stimulus and 4 s for them to remove the phoneme. Subjects' percent correct scores on the task were calculated and converted to age-regressed standardized z scores based on the results of the control group.

Print exposure. Stanovich and colleagues have argued that level of print exposure is a strong determinant of orthographic processing skills (Stanovich & West, 1989) and print exposure has been shown to load more strongly on an orthographic factor than on a phonological factor (Olson et al., 1994a). Therefore, we also included a print exposure test in the battery of validation measures, with the prediction that the surface dyslexic subgroup might obtain lower mean scores on this measure than the phonological dyslexic group. The task involved presenting children with the names of books and magazines, some of which were real and some of which were not, and asking them to identify those that they knew. Raw scores (percent hits-percent misses) were converted into standardized z scores based on the performance of the control group.

Word reading error types. The final validation measures examined the types of error subjects tended to make when they read a word on the TLWRT incorrectly. Specifically, subjects were scored on the percentage of phonetically correct responses or *regularizations* that they made (e.g., saying "sade" for *said*) and the percentage of other-word responses or *lexicalizations* that they produced (e.g., saying "sad" for *said*). It would be expected that poor readers relying on a phonological strategy might make more regularization errors than those relying on an orthographic strategy, while the latter might be expected to make more lexicalization errors. Each score was calculated as a percentage of the total number of errors the subject made.

RESULTS

Creation of the Subtype Variable

Standardized z scores on the orthographic and phonological processing measures were obtained for each of the 967 children in the RD group. The mean z score on the phonological processing measure (nonword reading) for this group as a whole was -1.72 ($SD = .95$) and their mean on the orthographic processing
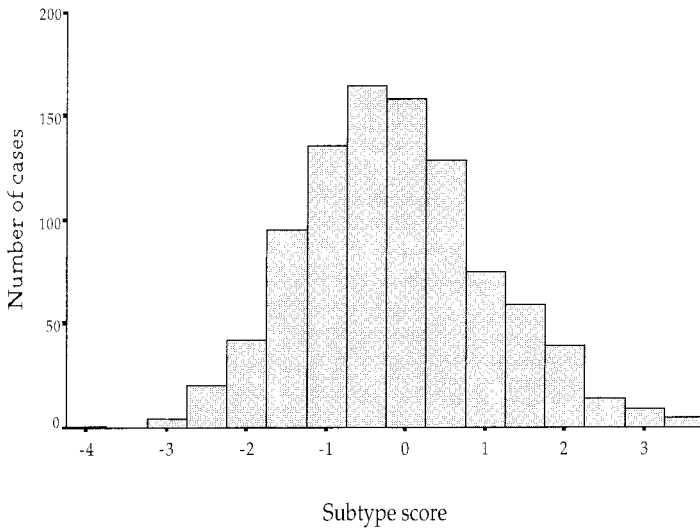
FIG. 1.   Distribution of scores of the reading-disabled subjects on the subtype variable.

measure (the proportion of strange words read incorrectly) was $-.90$ ($SD = 1.13$). Therefore, it is clear that poor readers as a group tended to be below average on both of these tasks. These scores were then used to create a subtype variable by subtracting the z score on the orthographic processing measure from the z score on the phonological processing measure for each subject.

Once this variable had been created, the subjects' overall reading level (as measured by the composite word recognition measure) was regressed out of the variable. The purpose of this was to ensure that any differences that subsequently resulted between subgroups chosen on the basis of this subtype were not due to general differences in the reading ability of the two groups. The variable was thus designed to reflect the relative level of subjects' orthographic and phonological skills, independently of their overall reading ability. The distribution of scores of the RD subjects on the subtype variable is presented in Fig. 1.

Selection of the RD Subgroups

Subjects at each end of the subtype distribution were chosen to represent the two reading subgroups of surface and phonological dyslexics. Specifically, those children whose subtype scores fell in the top third of the distribution (meaning that their standardized scores on the phonological processing measure were higher than their scores on the orthographic processing measure) were allocated to the surface dyslexic group, while children whose subtype scores fell into the bottom third of the distribution were allocated to the phonological dyslexic

TABLE 1
Characteristics of the Surface and Phonological Dyslexic Subgroups and Mean Scores
on the Subtyping Measures (with Standard Deviations in Parentheses)

	Surface dyslexics	Phonological dyslexics	t value	p
N	322	322		
Age	11.5 (2.7)	11.6 (2.4)	−0.4	>.05
IQ (Wechsler PIQ)	102 (11.5)	103 (10.8)	−1.1	>.05
Composite word recognition score (z)	−2.3 (1.1)	−2.3 (0.9)	−0.4	>.05
Subtype score (z)	1.3 (1.0)	−1.2 (0.5)	49.7	<.001
Nonword reading (z)	−1.3 (1.0)	−2.1 (0.7)	10.8	<.001
Strange word reading (z)	−1.8 (1.0)	0.0 (0.6)	−26.0	<.001

subgroup. There were 146 pairs where *both* twins fell into one or other of the two subgroups. Of these, 99 pairs (63MZ: 36DZ) fell into the same subgroup, while 57 pairs (22MZ; 35DZ) fell into different subgroups.

Descriptive statistics for the two subgroups can be found in Table 1. As would be expected, the subjects differed significantly on the defining orthographic and phonological processing tasks, but did not differ on age, nonverbal IQ, or overall word recognition level. The performance of subjects in these subgroups on the other reading and language measures was then examined to assess the validity of the group selections.

Validation Measures

Results for the two reading subgroups on the validation measures are presented in Table 2. As can be seen, the two subgroups differed significantly in the expected direction on all of the tasks used. The phonological dyslexic subjects performed more poorly than the surface dyslexics on the phonological choice

TABLE 2
Mean Scores on the Validation Measures by the Surface and Phonological Dyslexic Subgroups
(with Standard Deviations in Parentheses)

	Surface dyslexics	Phonological dyslexics	t value	p
Phonological choice (z)	−1.7 (1.2)	−2.1 (1.0)	3.1	<.01
Orthographic choice (z)	−1.4 (1.1)	−1.1 (1.1)	−2.8	<.01
Phoneme deletion (z)	−1.6 (1.7)	−2.2 (1.4)	3.2	<.01
Print exposure (z)	−1.3 (1.1)	−1.0 (1.0)	−2.2	<.05
Regularization errors (%)	23.5 (16.5)	18.2 (14.3)	4.3	<.001
Lexicalization errors (%)	28.2 (20.9)	32.4 (18.4)	−2.7	<.01

task and the phoneme deletion task and were also more likely to make lexicalization errors than the surface dyslexic subjects. The surface dyslexic groups, on the other hand, performed more poorly on the orthographic choice task and were more likely to make regularization errors than the phonological dyslexic group. The surface dyslexic group also performed more poorly on the print exposure measure than the phonological dyslexic group, although this result only just reached significance. We will take up this issue further in the Discussion.

Therefore, it appears that subgroups of poor readers can be identified who approach the task of reading words somewhat differently. The subjects in the present sample who fell at one end of the subtype dimension appeared to be more reliant on phonological reading skills, while those at the other end placed more emphasis on orthographic processing. Consistent with the findings of Manis et al. (1996) and Stanovich et al. (1997), the differences between these subgroups were apparent, not only in their performance on irregular word and nonword reading aloud tasks but also in their performance on related reading and language measures. With some validation of the differences between the subgroups having been obtained, we moved on to conducting analyses of genetic influences on the reading deficits in the two subgroups.

Genetic Analyses

Behavioral-genetic analyses of twin data often take the form of comparisons between covariance matrices for MZ and DZ twins. These kinds of analyses focus on individual differences across the general population and thus are not specifically designed to assess the heritability of *deviant* group membership. Therefore, a different form of behavioral-genetic analysis, which was designed to examine the heritability of group membership when twins were selected from the extreme low end of the normal distribution on the variable of interest (in this instance, reading ability), was adopted in the present study.

In this analysis, the twin of a pair who is deviant enough on the reading dimension to be called reading disabled is called the *proband* and the other member of the pair is called the *cotwin.* The heritability of the probands' group membership is then assessed by comparing the amount of regression toward the mean for the MZ and DZ cotwins. The logic is as follows: If the probands' group reading deficit was entirely due to genetic influence (and there was no test error), MZ cotwins would show no regression to the normal population mean because they share 100% of their genes. DZ cotwins would regress halfway to the population mean on average, since they share half of their segregating genes. On the other hand, if the probands' reading deficit was entirely due to shared environment influence, both MZ twins and DZ twins would show no regression toward the population mean. Therefore, by comparing relative MZ and DZ cotwin regression to the mean, estimates of the proportional genetic (h^2g) and shared environment (c^2g) influence on the probands' group membership in the low tail of the reading dimension can be obtained.

Defries and Fulker (1985) have developed a regression model for twin data to derive these estimates. The model is in the form

$$C = B_1P + B_2R + A,$$

where the cotwin's (C) score is predicted from the proband's (P) score, the coefficient of relationship (R; 1 for MZ twins and .5 for DZ twins) and a regression constant (A). B_1 provides an estimate of average MZ and DZ pair similarity, while B_2 estimates twice the difference between the means of the MZ and DZ cotwins. When the data are appropriately transformed, B_2 yields a direct estimate of the proportion of genetic influence on deviant-group membership (h^2g) (Defries & Fulker, 1988).[1] For this study, an extension of the basic model was also used, which allows differential heritability in relation to subtype variables to be tested (Defries & Fulker, 1985). The extension simply involves the inclusion in the model of the probands' subtype dimension (S) and the product of the coefficient of relationship and the subtype dimension (RS):

$$C = B_1P + B_2R + B_3S + B_4PS + B_5RS.$$

In this extended model, B_5 provides a test of the significance of differences in h^2g as a function of the subtype variable.

As described earlier, the reading-disabled sample or *proband* group in the present study consisted of those subjects who fell at least one standard deviation below age average on the composite single word reading task. In total, there were 592 pairs of twins (272 MZ; 320 DZ) in which at least one twin of the pair fell into this category. This total set of twin pairs formed the subject sample for the genetic analyses. In some cases, only one twin of a pair fell into the proband group, while in other cases both members did. If both members of a twin pair met the criteria for inclusion in the proband group, they were entered twice into the analysis, with the twin members exchanging proband and cotwin status. In these instances, both probands in a pair were assigned the mean subtype score for the two, so as to avoid having the same twin pair represented at different points along the subtype dimension. The standard errors for the estimates were based on the actual numbers of twin pairs in the analyses.[2]

We began by using the extended regression model to test the whole RD sample for the significance of the interaction between the heritability of the group deficit and the continuous subtype dimension (B_5). A significant result here would indicate that the heritability of reading deficits differs depending upon the position a proband is located at on the subtype dimension. Based on the earlier

---

[1] The transformation involves expressing the scores of the MZ and DZ twins as deviations from the unselected population mean and then dividing by the proband mean.

[2] This was achieved by multiplying the standard error by an adjusting factor of

$$\sqrt{N_{DE} - k - 1/N_{SE} - k - 1},$$

where N_{DE} = total number of double entered cases, N_{SE} total number of single entered cases, and k = the number of parameters in the equation.

work of Gayan et al. (1994), our expectation was that heritability estimates might tend to be higher toward the negative end of the subtype distribution, where deficits in phonological skills were relatively greater than those in orthographic processing than toward the positive end of the distribution, where the reverse was the case. The results yielded a B_5 estimate of .14 ($SE = .08$), which was significant in the predicted direction in a one-tailed test, $t(937) = 1.73$, $p < .05$. Thus, there was some evidence from the analysis of the whole RD sample that the heritability of word recognition deficits differed across the subtype dimension.

To explore this finding further, we used the basic model to obtain separate estimates of the heritability of word recognition deficits in the two extreme groups described previously. (The subjects included here actually differed slightly from those in the subgroups described earlier, due to the effect on group selection of averaging the subtype scores of double-entered twin pairs. However, the subgroups' results on the validation measures mirrored those of the earlier group selections). The phonological dyslexic subgroup consisted of 322 children who were at least one standard deviation below age average on the composite single word reading test and whose scores on the subtype dimension fell in the bottom third of the distribution. There were 208 twin pairs (97MZ; 111DZ) with at least one twin of the pair falling into this subgroup; this total set of twin pairs formed the sample used for the regression analysis. The estimated heritability component (h^2g) for this group was a highly significant .67 ($SE = .14$), $t(319) = 4.88$ $p < .001$. The estimated shared environmental influence (c^2g) was a smaller, but also significant, .27 ($SE = .13$), $t(319) = 2.01$, $p < .05$. Thus, there was evidence for both genetic and shared environmental components to word recognition deficits in the phonological dyslexic group, but the genetic component appeared to make the greater contribution of the two. The amount of regression toward the mean for both MZ and DZ cotwins in the phonological dyslexic group is represented in Fig. 2.

The surface dyslexic group consisted of 322 children who were also at least one standard deviation below average on the composite word reading test, but who in this case had scores on the subtype dimension in the top third of the distribution. The total sample used in the genetic analysis consisted of the 195 twin pairs (89 MZ; 106 DZ) with at least one twin of the pair falling into this subgroup. The analysis of genetic and environmental influences on this group's reading deficit revealed a small but significant estimated h^2g estimate of .31 ($SE = .13$), $t(319) = 2.41$, $p < .05$. The c^2g estimate of .63 ($SE = .13$) was highly significant, $t(319) = 4.72$, $p < .001$. Thus, the proportion of genetic and shared environmental influence on word recognition deficits in this subgroup differed markedly from that of the phonological dyslexics: In this case, the shared environmental component seemed to make the stronger contribution of the two. The regression toward the mean shown by MZ and DZ cotwins in the surface dyslexic group is also represented in Fig. 2.
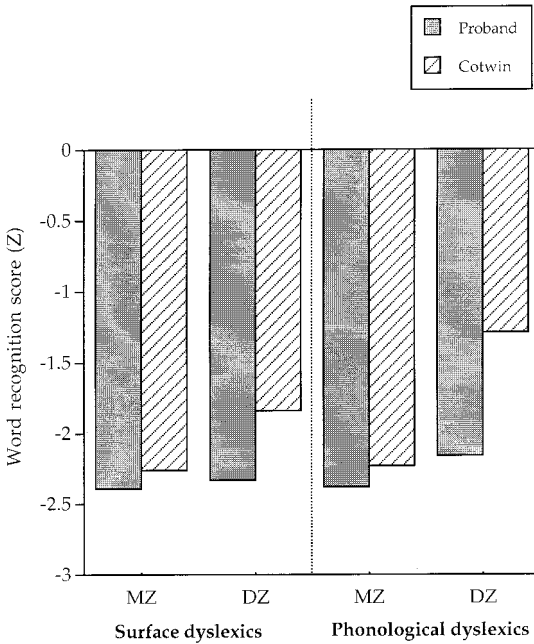
FIG. 2. Cotwin regression toward the mean for word recognition in the surface and phonological dyslexic subgroups.

Of course, selecting subgroups at the top and bottom thirds of the distribution represented an arbitrary cut-off. To confirm the result, we conducted the same analyses with subgroups selected from the top and bottom quarters and fifths of the distribution. This produced similar, if even more extreme, results: h^2gs for phonological dyslexics selected from the bottom quarter and bottom fifth of the distribution were .73 ($SE = .15$) and .78 ($SE = .18$), respectively, while c^2gs were .20 ($SE = .12$) and .13 ($SE = .16$), respectively. Surface dyslexics chosen from the top quarter and top fifth of the distribution had h^2gs of .32 ($SE = .14$) and .37 ($SE = .16$) and c^2gs of .63 ($SE = .12$) and .57 ($SE = .15$), respectively.

Thus, using these measures, there was some evidence for differential heritability of word recognition deficits across subtypes of poor readers. In a set of supplementary analyses, we attempted to specify more precisely the basis of these findings. First, we looked at the interaction between the separate components of the subtype dimension and the heritability of the group deficit in word recognition to determine which component of the subtype dimension was accounting for the most variance in the data. The subjects' standardized z scores on the phonological measure (nonword percent correct) and the orthographic measure (strange word reading) were entered in separate analyses into the extended regression model as subtype dimensions. Again, the subjects' overall scores on

the word recognition measure were regressed out of the subtype scores and the mean subtype score of a twin pair was used in those instances where both members of a pair fell into the proband group. The interaction between the phonological subtype and the heritability of the group deficit was not significant, $B_5 = -.08$ ($SE = .11$), $t(953) = .68$, $p > .05$. For the orthographic measure, the interaction was larger, $B_5 = .14$ ($SE = .09$), but still did not reach significance, $t(953) = 1.51$, $p > .05$. Thus, it would appear that, even though the strange word reading measure was the most significant source of the variation in the subtype scores, both measures contributed to the interaction between the subtype variable and the heritability of word recognition deficits.

Following this, we examined more closely the composite reading measure on which the proband group was originally selected. As discussed, this composite measure was derived in equal weights from subjects' scores on the TLWRT and the Word Recognition subtest of the PIAT. We therefore selected two new proband groups based on subjects' scores on the TLWRT and PIAT measures *separately* (in both cases, the criterion for selection was one standard deviation below age average on the measure used). The subtype dimension was calculated in the same way as previously (nonword percent correct minus strange word percent correct) and the extended model was then used to examine the interaction between the heritability of the group deficit on each of the tasks and the subtype dimension. In both cases, the interaction was in the same direction as was found previously, but in neither case did the term reach significance. For the TLWRT proband group (N twin pairs $= 566$; 261 MZ; 305 DZ), $B_5 = -.12$ ($SE = .08$), $t(910) = 1.50$, $p > .05$. For the PIAT proband group (N twin pairs $= 568$; 255 MZ; 313 DZ), $B_5 = -.11$ ($SE = .08$), $t(916) = 1.39$, $p > .05$. Thus, although the heritability of deficits on both measures of single word recognition showed a tendency toward an interaction with the subtype dimension, the composite measure produced the strongest interaction.

DISCUSSION

The present results provide further evidence that it is possible to identify subgroups of children who demonstrate different patterns of reading disability. Following from the work of Manis et al. (1996) and Stanovich et al. (1997), we selected poor readers who fell at the two extreme ends of the distribution of our subtype dimension and then compared them on a range of reading and language measures. The results were consistent with the hypothesis that subjects in one group were relying primarily on a strategy of grapheme–phoneme conversion, while those in the other group were relying more on the use of word-specific information. This was evident in the types of errors that they made when reading aloud, as well as in their performance on other tasks designed to measure orthographic and phonological processing. Specifically, the phonological dys-lexic subjects had more difficulty than the surface dyslexics determining if

written nonsense words sounded like words and removing single phonemes from spoken words. They were also more likely to produce regularization errors when reading aloud single words. The surface dyslexics had greater difficulty distinguishing words from pseudohomophones and were less likely than the phonological dyslexics to produce lexicalization errors when attempting to read single words aloud.

We wish to emphasize that we do not see the poor readers in the surface and phonological dyslexic subgroups as representing distinct "subtypes," who are qualitatively different from other dyslexics, but rather simply as those subjects who fall at the extreme ends of the distributions of performance on orthographic and phonological processing (Olson et al., 1985). Based on an assumption that these processes can develop independently, at least to some degree, we interpret the surface dyslexic pattern in terms of relatively better development of phonological than orthographic skills and the phonological dyslexic pattern as the reverse. The identification of these subgroups is seen as being of interest for two main reasons. First, as Castles and Coltheart (1993) have argued, it provides evidence that irregular word reading and nonword reading can be dissociated and, as such, has presented a challenge to models of the reading system that do not distinguish between processes involved in performing these tasks (Seidenberg & McClelland, 1989). Second, since the poor readers in these two extreme groups have relatively strong deficits in either orthographic or phonological skills, they represent useful subject samples for exploring the bases of such deficits.

The behavioral genetic analyses reported here represent one approach to exploring the etiology of deficits in word recognition depending on subjects' position on a phonological-surface subtype dimension. With regard to phonological dyslexia, the results are suggestive of a strong genetic component to the disorder: Approximately two thirds of the reading deficit of the phonological dyslexic probands was estimated to be due to heritable influences. In contrast, only about one quarter of the reading deficit in these subjects was found to be attributable to the influence of shared environment. This finding is consistent with previous work by Olson and colleagues, which has indicated substantial heritability of phonological deficits in poor readers who have not been differentiated by subtype (Olson et al., 1989, 1994a). Also, given the large amount of evidence that now exists for a relationship between impairments in nonword reading and deficits in the analysis of the phonemic structure of spoken words (see e.g., Bowey et al., 1992; Bruck, 1992; Manis et al., 1993; Olson et al., 1989; Pennington et al., 1990), the results provide converging support for the prevailing view that phonological dyslexia is associated with a fundamental, and possibly inherited, spoken language processing deficit.

The results for the surface dyslexic subjects were notably different from those of the phonological dyslexics. Although there was evidence for a small but significant genetic component to the reading deficit in these subjects, the influ-

ence of shared environment was found to be much greater: More than 60% of the reading deficit in the surface dyslexic probands was estimated to be due to shared environmental influences.

It should be noted, however, that the orthographic processing measure used in the present study was somewhat limited, consisting only of the relatively small number of strange words read by the subjects. Although this measure successfully discriminated between the surface and phonological subgroups, it would add weight to the present findings if some validation of the results using other orthographic measures could be found. Some preliminary results reported by Gayan et al. (1997), drawing from the same database as used here, would seem to provide such a validation. Gayan et al. created a subtype dimension based on the subjects' performance on the same oral nonword reading task as used here and the orthographic choice task of Olson et al. (1989, 1994b). Phonological and surface dyslexic subgroups were then selected from the top and bottom thirds of the distribution. The heritability of the group deficits in word recognition (h^2g) were found to be .35 and .53 for the surface and phonological dyslexic groups respectively, while the c^2gs were .61 and .38, respectively. Thus, although the difference between the two groups in the degree of genetic influence on word recognition deficits was not as strong as was found in the present study, the trend of the results was clearly in the same direction.

How, then, do these behavioral genetic data accord with the differing accounts of surface dyslexia outlined earlier? We noted in the introduction that the Manis et al. (1996) hypothesis of a general resource deficit may be consistent with both genetic and environmental influences. Similarly, a weakness in the lexical route of the dual-route cascading model of Coltheart et al. (1993) is compatible with both genetic and environmental etiologies. However, the differential genetic etiology that was found for the surface and phonological dyslexic subtypes does support the partial independence in development of the lexical and nonlexical routes proposed in the dual-route model.

The present behavioral-genetic results do seem consistent with a third view of surface dyslexia: Stanovich et al. (1997) propose that the surface dyslexic pattern reflects a mild phonological deficit which has been compounded by lack of exposure to reading materials. This account is consistent with the present finding of a small but significant genetic component to the reading deficit and a much larger environmental component. It also accords with the results for the surface dyslexics on phonological processing tasks: They were indeed impaired on these tasks, relative to normal readers, but not to the degree that the phonological dyslexics were. Therefore, to this extent, Stanovich et al.'s theory would appear to provide a better account of the present set of behavioral genetic and phenotypic data than that of Manis et al.

However, there are other aspects of the data which are less consistent with Stanovich et al.'s (1997) account. A clear prediction from this theory would

be that surface dyslexics should show markedly lower scores on measures of text exposure than phonological dyslexics. Although the surface dyslexics did score slightly lower on the print exposure measure used in the present study than did the phonological dyslexics ($z = -1.3$ vs $z = -1.0$, respectively), the difference between the two groups was only marginally significant. This slightly reduced performance by the surface dyslexics would seem insufficient to account for their substantial deficits on measures of orthographic processing. In addition, given that the phonological dyslexics also scored on average one standard deviation below the control subjects on the print exposure measure, it is unclear why these subjects' reading pattern did not more closely resemble that of the surface dyslexics. Two possible explanations of this result can be suggested. One is that deficits in text exposure will result in a phonological dyslexic pattern in poor readers with a severe phonological deficit but in a surface dyslexic pattern in those with a milder phonological deficit. The other possibility is that the present result can be attributed to psychometric limitations of the measure used. As Olson et al. (1994a) have previously noted, the print exposure measure may have large error variance due to guessing, and familiarity with the specific titles in the measure may be only loosely tied to the amount of exposure to books with other titles. Also, the book titles used may not have been appropriate for the full age range of subjects in the sample.

While an environmental deficit in print exposure is a plausible account of surface dyslexia, other environmental factors are equally plausible. We noted in the introduction that intense instruction in phonological processing can result in a surface dyslexic profile (Olson et al., 1997). We also noted the possibility that environmental factors such as lead exposure could lead to a constitutional deficit in general processing resources that is consistent with the account of Manis et al. (1996). Prenatal factors such as exposure to alcohol, flu viruses, and other teratogens might also constitute environmental influences. Of course, it is also possible that several types of environmental factors operate to different degrees in individual children with a surface dyslexic profile. One point is clear from the behavior-genetic analyses of environmental influences on reading deficits in twins with surface dyslexia: The environmental influence is largely shared by the twins in a pair instead of being unique for each twin.

In conclusion, much more work is needed to establish the bases of orthographic and phonological reading deficits. Although there is now considerable evidence that specific patterns of reading disorder can be identified, the precise nature of the impairments which underlie these patterns remains relatively underspecified, particularly in the case of surface dyslexia. The behavioral genetic analyses reported here represent one means of addressing these issues. It is to be hoped that this work, together with converging evidence from other sources, will lead to an increased understanding of the nature and causes of specific reading impairments.

APPENDIX

Strange Items from the TLWRT

Screener Items

1. foreign	3. grotesque
2. aerial	4. baroque

Test Items

1. oh	19. feudal
2. eye	20. hymn
3. ocean	21. nasal
4. fruit	22. lieutenant
5. juice	23. psychiatrist
6. clothes	24. isle
7. neighbor	25. chauffeur
8. daughter	26. beseige
9. cocoa	27. acreage
10. muscle	28. colleague
11. coyote	29. hieroglyphics
12. bruise	30. chaos
13. neuter	31. cyst
14. league	32. suede
15. yacht	33. mitigate
16. gnat	34. brusque
17. camouflage	35. eulogy
18. unique	36. demagogue

REFERENCES

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three typical reading-spelling patterns. *Developmental Medicine and Child Neurology,* **15,** 663–687.

Bowey, J. A., Cain, M. T., & Ryan, S. M. (1992). A reading level design study of phonological skills underlying fourth-grade children's word reading difficulties. *Child Development,* **63,** 999–1011.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature,* **271,** 746–747.

Bruce, D. J. (1964). The analysis of word sounds by young children. *British Journal of Psychology,* **34,** 158–170.

Bruck, M. (1992). Persistence of dyslexics' phonological awareness deficits. *Developmental Psychology,* **28,** 874–886.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition,* **47,** 149–180.

Castles, A., & Holmes, V. M. (1996). Subtypes of developmental dyslexia and lexical acquisition. *Australian Journal of Psychology,* **48,** 130–135.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing.* London: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual route and parallel-distributed-processing approaches. *Psychological Review,* **100,** 589–608.

Defries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics,* **15,** 467–473.

Defries, J. C., & Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae: Twin Research,* **37,** 205–216.

Fletcher, J. M., & Morris, R. (1986). Classification of disabled learners: Beyond exclusionary definitions. In S. J. Ceci (Eds.), *Handbook of social and neuropsychological aspects of learning disabilities* (Vol. 2, pp. 55–80). Hillsdale NJ: Erlbaum.

Gayan, J., Datta, H. E., Castles, A. E., & Olson, R. K. (1997). *The etiology of group deficits in word decoding across levels of phonological decoding and orthographic coding.* Paper presented at the Annual Meeting of the Society for the Scientific Study of Reading, Chicago, March 22–24.

Gayan, J., Forsberg, H., & Olson, R. K. (1994). Genetic influences on subtypes of dyslexia. *Behavioral Genetics,* **24,** 513.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education,* **6,** 8–17.

Manis, F. R., Custudio, R., & Szezulski, P. A. (1993). Development of phonological and orthographic skills: A 2-year longitudinal study of dyslexic children. *Journal of Experimental Child Psychology,* **56,** 64–86.

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Peterson, A. (1996). On the bases of two subtypes of developmental dyslexia. *Cognition,* **58,** 157–195.

Morton, J., & Patterson, K. E. (1980). A new attempt at an interpretation or an attempt at a new interpretation. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Ed.), *Deep dyslexia.* London: Routledge and Kegan Paul.

Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual and developmental differences in reading disability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (pp. 1–64). New York: Academic Press.

Olson, R. K., Forsberg, H., & Wise, B. (1994a). Genes, environment and the development of orthographic skills. In V. Berninger (Ed.), *Varieties of orthographic knowledge: Theoretical and developmental issues* (Vol. 1, pp. 27–71). Dordrecht, The Netherlands: Kluwer Academic.

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994b). Measurement of word recognition, orthographic and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243–277). Baltimore: Paul H. Brooks.

Olson, R., Wise, B., Conners, F., Rack, J., & Fulker, D. (1989). Specific deficits in component reading and language skills: Genetic and environmental influences. *Journal of Learning Disabilities,* **22,** 339–348.

Olson, R., Wise, B. W., Ring, J., & Johnson, M. (1997). Computer-based remedial training in phoneme awareness and phonological decoding: Effects on the post-training development on word recognition. *Scientific Studies of Reading,* **1,** 235–253.

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,* **57,** 1001–1013.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review,* **103,** 56–115.

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.

Rosner, J., & Simon, D. (1971). The Auditory Analysis Test: An initial report. *Journal of Learning Disabilities,* **4,** 384–392.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review,* **96,** 523–568.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior,* **23,** 383–404.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly,* **24,** 402–433.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology,* **89,** 114–127.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). The development of reading related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology,* **30,** 73–87.