Journal of Educational Psychology
1996, Vol. 88, No. 1, 3–17

Developmental Lag Versus Deficit Models of Reading Disability: A Longitudinal, Individual Growth Curves Analysis

David J. Francis
University of Houston

Sally E. Shaywitz
Yale University School of Medicine

Karla K. Stuebing
University of Texas—Houston Medical School

Bennett A. Shaywitz
Yale University School of Medicine

Jack M. Fletcher
University of Texas—Houston Medical School

Individual growth curves were used to test whether the development of children with reading disabilities is best characterized by models of developmental lag or developmental deficit. Developmental changes in reading ability were modeled by using 9 yearly longitudinal assessments of a sample of 403 children classified into three groups representing (a) deficient reading achievement relative to IQ expectations (RD-D), (b) deficient reading achievement consistent with IQ expectations (LA), and (c) no reading deficiency (NRI). Using a model of quadratic growth to a plateau, the age and level at which reading scores plateaued were estimated for each child. Reading-disabled children differed on average from nondisabled children in the level but not in the age at which reading skills plateaued. The RD-D and LA groups did not differ in reading plateau or age at plateau. The subgroup of RD-D children scoring below the 25th percentile in reading differed from LA children only in reading plateau. Results suggest that the developmental course of reading skills in children with reading disability is best characterized by deficit as opposed to lag models. In addition, no support for the validity of classifications of reading disability based on IQ discrepancies was apparent.

Despite 30 years of research, it remains unclear whether children who vary in reading abilities are best characterized by models involving developmental lags or deficits (Fletcher, 1981; Stanovich, Nathan, & Vala-Rossi, 1986). The deficit model assumes that children fail to read proficiently because of the absence of a skill that never develops sufficiently (Cromer, 1970). In contrast, the lag model is based on the hypothesis that children who differ in reading ability vary only in the rate at which cognitive skills develop, so the skill will emerge over time.

The issue of developmental lag versus deficit has important implications for theoretical models of children with reading disability. The models were initially formulated to represent different hypotheses about the neurobiology of reading disability, with some investigators specifically hypothesizing that reading disability represented a lag in the maturation of the brain (Satz, Taylor, Friel, & Fletcher, 1978). These applications of lag models were tested in longitudinal studies of a variety of skills in disabled readers that were thought to index brain function. The maturational lag hypothesis predicted that poor readers would "catch up" to their peers on these skills as the brain matured. In contrast, a deficit model hypothesized that brain organization in children with reading disability was simply disordered so that no catching up would occur (Rourke, 1976).

More recently, the issue of lag versus deficit has been approached in terms of the implications of these models for skill development and intervention. In these studies, this issue was approached by using a reading-level match design that compares cognitive development between older, poor readers and younger, skilled readers who are matched on reading ability. A group difference is usually interpreted as an example of a skill that has significant implications for the reading disability, warranting intervention. In contrast, if no

David J. Francis, Department of Psychology, University of Houston; Sally E. Shaywitz, Department of Pediatrics, Yale University School of Medicine; Karla K. Stuebing and Jack M. Fletcher, Department of Pediatrics, University of Texas—Houston Medical School; Bennett A. Shaywitz, Departments of Pediatrics and Neurology, Yale University School of Medicine.

This research was supported in part by National Institute of Child Health and Human Development Grants PO1 HD21888, *Psycholinguistic and Biological Mechanisms in Dyslexia,* P50 25802, *Center for Learning and Attention Disorders,* HD 28172, *Detecting Reading Problems by Modeling Individual Growth,* and HD 30995 *Early Intervention for Children with Reading Problems.*

The assistance of John Holahan and Abigail Shneider is gratefully acknowledged. We thank the parents, children, and educators whose participation in the Connecticut Longitudinal Study has contributed so much to furthering our understanding of reading and reading problems. We also acknowledge the long-term intellectual contribution of Paul Satz to our thinking about the developmental course of reading disability.

Correspondence concerning this article should be addressed to David J. Francis, Department of Psychology, University of Houston, 4800 Calhoun, Houston, Texas 77204-5341. Electronic mail may be sent via Internet to dfrancis@uh.edu.

difference emerges, the skill is assumed to be merely lagging behind in the poor reader who will eventually catch up, even in the absence of intervention. Obviously, if the development of reading skills was shown to fit a lag model in which children who read poorly earlier in development eventually catch up to their peers, it could be argued that intervention and identification would not be necessary. In this case, policymakers would surely argue that intervention is unnecessary unless it could be shown that intervention reduces the time required for skills to catch up and that there is some benefit to catching up sooner as opposed to later.

There are several versions of the lag model. A strong version hypothesizes that children who are poor readers differ from normal readers in the rate at which cognitive skills develop, but they eventually catch up to children who are comparable in age but who exhibit better initial reading. Consequently, from a strong version of the lag model, there should be no differences between disabled and nondisabled readers in the end point development of reading and other cognitive skills. One way for this result to be obtained is for the rate of development in poorer readers to exceed the rate of development in more skilled readers at some point during the developmental process (Fletcher, 1981; Satz, Fletcher, Clarke, & Morris, 1982; Stanovich et al., 1986; Stanovich, Nathan, & Zolman, 1988). This pattern of results is also consistent with the notion of protracted development in impaired readers.

A weak version of the lag model hypothesizes that poor readers vary in the rate of development of cognitive skills, but it does not address the issue of end point development of reading ability. This version simply predicts no differences in cognitive ability between older, poor readers and younger, skilled readers who are comparable in reading level.

The weak version has been evaluated with reading-level match designs. Comparisons of older, poor readers and younger, skilled readers matched on reading level on a variety of reading and other cognitive skills frequently show no differences in the development of some cognitive skills, particularly outside of the domain of phonological language. However, older, poor readers invariably show poorer pseudoword identification skills and problems with phonological language, particularly when reading is measured with word recognition tasks (Backman, Bruck, Hebert, & Seidenberg, 1984; Beech & Harding, 1984; Guthrie, 1973; Guthrie & Tyler, 1975; Olson, Kliegl, Davidson, & Foltz, 1985; Stanovich et al., 1986). The null results for certain skills, while presenting interpretive problems (Backman, Mamen, & Ferguson, 1984; Jackson & Butterfield, 1989; Stanovich & Siegel, 1994; Vellutino & Scanlon, 1989), are generally accepted as supporting a lag model, rarely specifying whether a strong or weak version is being evaluated.

More recently, Stanovich and Siegel (1994) developed a regression-based approach that avoided some of the interpretive problems apparent when older, poor readers are matched with younger, good readers in reading ability. Evaluating a broad range of hypotheses and outcomes, Stanovich and Siegel (1994) found little support in this cross-sectional study for the developmental lag hypothesis, concluding that "a developmental lag model in its strongest form does not fit the present results very well" (p. 48). However, these authors did not actually test a strong version of the lag model, as defined in this article, because children were evaluated only once, so end point development of reading skills could not be addressed.

Like Stanovich and Siegel's study (1994), most of the reading-level studies are cross-sectional and include only a single occasion of measurement. With only one time point, a reading-level design cannot address whether disabled children catch up in reading ability relative to nondisabled children. In the few reading-level design studies with a longitudinal follow-up, researchers have found smaller cognitive differences between skilled and less skilled readers relative to the large differences in cognitive functions apparent in chronological-age comparisons. However, when these children were assessed later in development, the differences were larger than in the initial assessment (Guthrie & Seifert, 1977; Stanovich et al., 1988). This finding implies that poor readers are not characterized by an accelerated rate of development at later periods. Hence, the null results found in cross-sectional reading-level designs may not inform researchers about differential developmental rates.

When same-aged children differing in reading ability are compared (i.e., chronological age design), large differences in the cognitive functioning of good and poor readers are common. Some skills with a rapid rate of development between 5 and 7 years show evidence of catching up by poor readers, such as the ability to match geometric shapes or identify a finger that has been touched by the examiner outside of the child's view (Rourke, 1976). However, there is little evidence from longitudinal studies that the ultimate level of development of either the cognitive or the reading skills of a less skilled reader ever approximates that of a comparably aged skilled reader (Rourke, 1976; Satz & Fletcher, 1980; Satz et al., 1982). Hence, these results are more consistent with a deficit model.

With the exception of Stanovich and Siegel (1994), much of the evidence supporting a developmental lag model has come from studies of skilled and less skilled readers—not of children identified as disabled in reading on the basis of statistical criteria, such as a discrepancy between IQ and reading test performance. For these children, it has been hypothesized that lag models more appropriately characterize children who have reading abilities consistent with intelligence test scores (i.e., garden-variety or general backward readers; Gough & Tumner, 1986; Rutter & Yule, 1975; Stanovich et al., 1988; Stanovich & Siegel, 1994). It has also been hypothesized that lag models may fit less well for children with "specific" reading disabilities whose reading scores are well below expectations for their IQ scores. Stanovich et al. hypothesized that lag models may be more appropriate for children who are garden-variety poor readers (i.e., low achievers) or for comparisons of skilled and less skilled readers—not for children with specific reading disabilities.

Although these hypotheses have not been confirmed, finding that children with specific reading disabilities and garden-variety reading disabilities differ in models of growth would provide validation of the common differentiation of children with general reading backwardness and specific reading retardation set forth by Rutter and Yule (1975). In recent studies researchers have not found differences in cognitive skills between these two groups of disabled readers by using either chronological-age (Fletcher et al., 1994) or reading-level (Stanovich & Siegel, 1994) designs. However, in a longitudinal follow-up, Rutter and Yule also found that general backward readers showed more rapid development of reading skills than children with specific reading retardation. To determine whether the developmental course varies between these types of disabled readers requires a longitudinal design with multiple follow-up assessments to eliminate problems of regression to the mean. If it was shown that the developmental course was different, such findings would support the commonly held view that these two reading groups represent fundamentally different types of reading disability.

Few studies addressing either a strong or weak version of the lag hypothesis have used a longitudinal design. The only large longitudinal follow-up of children with reading disabilities provided no evidence of catching up (or comparable rates of development) on cognitive measures evaluated at 5, 8, and 11 years (Fletcher, Satz, & Morris, 1984; Satz et al., 1978). However, this study did not provide comparable test-based assessments of reading ability over time. Although teacher ratings of the child's reading level were obtained yearly, performance-based assessments of actual word reading ability were only obtained in Grades 2 and 5. More important, different word recognition tests were used at the two follow-ups, which complicates any inferences about growth in reading skill. In fact, few studies of children with reading disabilities defined according to statistical criteria have assessed actual reading abilities longitudinally with more than two time points. Regardless of whether a chronological-age or reading-level design was used, such a study could directly test both strong and weak versions of the developmental lag hypothesis of reading disability.

In the present study we used individual growth curves analysis to directly evaluate the developmental lag hypothesis (Bryk & Raudenbush, 1987; Burchinal & Appelbaum, 1991; Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1994; Rogosa, Brandt, & Zimowski, 1982). Individual growth curves permit quantification and description of change at the individual level. A large sample of children was available in which reading skills had been measured annually in Grades 1–9 (S. E. Shaywitz, Shaywitz, Fletcher, & Escobar, 1990), making it possible to evaluate the lag and deficit hypotheses with longitudinal data. This sample has been used for three previous longitudinal studies. S. E. Shaywitz, Escobar, Shaywitz, Fletcher, and Makuch (1992) used data from Grades 1 through 5 and models of the multivariate normal distribution to evaluate the hypothesis that reading ability had a bimodal distribution. Results of this study demonstrated that the joint distribution of reading achievement and intelligence is bivariate normal at the three ages studied. Moreover, this study demonstrated that the classification of children as disabled or nondisabled on the basis of a discrepancy between IQ and achievement was more stable when classification was based on Grade 3 performance rather than on Grade 1 performance. In a more recent study, B. A. Shaywitz et al. (1995) used regression models to evaluate the Matthew effect hypothesis (i.e., the notion that reading skills and intelligence have cumulative effects on development such that the differences between good and poor performers get larger with age). This study was restricted to data collected in Grades 1–6 and did not distinguish between disabled and nondisabled readers. Rather, B. A. Shaywitz et al. (1995) investigated the overall correlation between level of performance and change in performance by using linear regression models for intelligence and reading. These authors concluded that there is evidence of a small Matthew effect for intelligence but that there is no evidence of such an effect for reading. Finally, Ishwaran et al. (1995) used structural equation models to evaluate the psychometric properties, including the longitudinal stability, of IQ–Achievement discrepancy scores. In addition, these authors examined the temporal stability of IQ and achievement scores. This study did not distinguish between disabled and nondisabled readers but did show greater reliability for IQ–Achievement discrepancies at the construct level beginning at Grade 3, which lends support to the observations of S. E. Shaywitz et al. (1992) that disability classifications based on observed scores are less stable in Grades 1 and 2 than classifications in Grade 3 and later. Although each of these studies was longitudinal in nature and the same sample of children was used, the hypotheses evaluated are quite distinct from the hypotheses addressed in the present article.

For this study, hypotheses of lag and deficit were operationalized in a model for the individual development of reading ability. These models were then used to compare children with reading disabilities defined by specific psychometric–statistical criteria, including children with reading scores consistent with intelligence test scores (low achievement) and children with reading scores below expectation based on intelligence test scores (discrepancy). Although Stanovich and Siegel (1994) did not find evidence of differences between low-achieving readers and discrepant disabled readers, their study was not longitudinal. In addition, the validity of conclusions concerning developmental lags in reading-level designs and the regression approach utilized in Stanovich and Siegel was dependent on change being linear throughout the age range studied. If changes in reading and other cognitive skills are nonlinear, differences in linear rates of change may not fully characterize differences in development among reading groups. The validity of inferences regarding the lag and deficit hypotheses optimally requires characterization of change at the individual and group level.

Williamson, Appelbaum, and Epanchin (1991) used linear models to evaluate change over time in the development of reading and math abilities. In contrast, we hypothesized that changes in reading ability would be nonlinear, showing an initial pattern of rapid acquisition with subsequent slowing in the rate of change. Consistent with the view that reading disabilities occur on a continuum and are a variation on normal development (S. E. Shaywitz et al., 1992; Stanovich, 1988), we did not expect to find that different models of change would characterize disabled readers who met discrepant and low-achievement definitions. However, if change is nonlinear and these groups differ in the rate of change, it is possible that the hypothesis put forward, but not supported by Stanovich and Siegel (1994), would be endorsed. Under this hypothesis, a lag model would fit the low-achieving readers, and a deficit model would fit the discrepant readers. Because Rutter and Yule (1975) showed that low-achieving readers had a better prognosis for the development of reading skills than the discrepant readers, the low achievers should show more catching up in reading ability than the discrepant group. Such findings would support the validity of this two-group classification of children with reading disabilities.

Method

Participants

The children for this study ($N = 403$) represented participants in the Connecticut Longitudinal Study (CLS; S. E. Shaywitz et al., 1990), a cohort of Connecticut children identified in kindergarten and assessed on a variety of cognitive, behavioral, and academic skills in Grades 1–9 (B. A. Shaywitz et al., 1995; S. E. Shaywitz et al., 1992; S. E. Shaywitz et al., 1990). Children included in the CLS were selected by recruiting all children in two randomly chosen kindergarten classes within each of 12 communities representative of the sociodemographic characteristics of the state of Connecticut as a whole. Each kindergarten class in these 12 towns had an equal probability of selection. All of the children in the 24 selected classes were invited to participate in the study. The final sample was clearly representative of the Connecticut kindergartners in 1983, when the CLS began. Exclusionary criteria were limited to significant sensory impairment and to serious psychiatric disorders. The initial sample ($N = 445$) represented 97% of the selected sample, including 235 girls (53%) and 210 boys (47%). The children included 375 Whites (84%), 50 African Americans (11%), and 20 children in other races (5%).

For this study, the Rasch-scaled reading-cluster score from the *Woodcock–Johnson Psychoeducational Test Battery* (Woodcock & Johnson, 1977) in Grades 1–9 was used as the primary measure of reading ability for longitudinal analysis. The reading-cluster score is a composite of the Word Identification, Word Attack, and Passage Comprehension subtests. We selected the cluster score because it has higher reliability than the individual subtests. Several scales for the reading-cluster score are provided in the Woodcock–Johnson scoring manual. Most readers are familiar with age-standardized and grade-equivalent scales. The Rasch-scaled score reported for the reading-cluster score is a transformation of the number correct for each subtest that yields a score with interval scale properties and a constant metric. The transformation is such that a score of 500 corresponds to the average performance level of beginning fifth graders. Its interval scale and constant metric properties make the Rasch-scaled score ideal for longitudinal studies of individual growth (Bryk & Raudenbush, 1987). For more technical information on item response models and Rasch scores in general, readers should see Hambleton and Swaminathan (1985) and Hambleton, Swaminathan, and Rogers (1991). The Woodcock–Johnson scoring manual provides information about the specific score transformation used in that test and about the general properties of the score.

In addition to measuring reading achievement, the Wechsler Intelligence Scale for Children—Revised (WISC–R; Wechsler, 1974) was obtained in Grades 1, 3, 5, 7, and 9. Determination of disability in reading was made at Grade 3, the earliest time point in this sample at which such problems can be identified with stability over time (S. E. Shaywitz et al., 1992). Reading disability could not be defined at Grade 1 or 2 because the WISC–R was not obtained in Grade 2 and identification of children in Grade 1 with these definitions overidentifies children who have not had adequate educational exposure to reading and underidentifies children who demonstrate deficits in phonological language. This is why previous studies have shown that identification of children in Grade 1 as disabled based on IQ and reading tests is unstable (Share & Silva, 1986; S. E. Shaywitz et al., 1992). For their data to be used in the current study, children were required to have a WISC–R in Grade 3, which resulted in the loss of 21 children. Consistent with traditional exclusionary criteria, Grade 3 WISC–R full-scale IQ scores were required to be 80 or higher, which resulted in the loss of 10 children. Finally, children were required to have at least six reading assessments, resulting in the loss of 1 more children.

Two groups of disabled readers were identified by using regression-adjusted discrepancy or low-achievement criteria (Fletcher et al., 1994): (a) reading disabled–discrepant (RD-D, $n = 32$) made up of children whose actual reading achievement score at Grade was at least 1.5 standard errors (*SE*) below the score predicted from the child's IQ, and (b) reading disabled–low achievers (LA, $n = 37$) made up of children who did not meet the regression criterion for group RD-D (i.e., children who did not show a 1.5 *SE* difference between actual and predicted achievement) but who had age-standardized reading cluster scores below the 25th percentile. It should be pointed out that our approach to classification gave first priority to the discrepancy between IQ and achievement. Any child meeting the regression criterion was classified as RD-D regardless of his or her level of achievement. For any child whose achievement score fell below the 25th percentile, that child was classified as RD-D if he or she met the regression criterion and as LA if she or he did not meet the regression criterion. Of the remaining sample, 334 children met neither definition and were considered not reading impaired (NRI). These criteria represent an operationalization of standard methods for identifying children as reading disabled in the public schools and in research studies (Frankenberger & Fronzaglio, 1991).

Table 1 presents mean Grade 3 WISC–R (Wechsler, 1974) full-scale IQ and Woodcock–Johnson (1977) reading-cluster age standardized scores for the three groups of children. The values in Table 1 show a familiar pattern of IQ and reading scores when regression-based discrepancy and low-achievement cutoff definitions are jointly applied in the same population (see Figure in Fletcher et al., 1994). On average, children in the RD-D group had a higher full-scale IQ, $t(67) = -6.23$, $p < .0001$. Although, on average, reading achievement scores are not significantly different between the two disability groups, which is obvious from Table 1, the RD-D children have lower reading scores than the LA children at any given level of IQ. This pattern emerges because, by definition, the IQ–Achievement discrepancy rule forms an upper bound for achievement of RD-D children at every level of IQ. Children whose achievement scores lie below this boundary are classified as discrepant; children whose achievement scores lie above this line are either classified as LA or NRI. Thus, the RD-D group consists of the lowest achievers at any given level of IQ. In a majority of cases, these children also fall below the cutoff for

Table 1
Distribution of WISC–R Full-Scale IQ and Woodcock–Johnson (1977) Reading-Cluster Scores for Three Reading Groups at Grade 3

	WISC–R full-scale IQ			Reading cluster (age standardized)	
Reading group	*n*	*M*	*SD*	*M*	*SD*
Not reading impaired	334	113.0	13.4	108.3	10.7
Reading disabled–discrepant	32	110.5	14.4	85.2	8.6
Low achieving	37	93.8	7.1	85.3	2.6

Note. As described in the text, 19 of 32 reading disabled–discrepant children were also low achieving but were classified only as discrepant unless explicitly stated. WISC–R = Wechsler Intelligence Scale for Children—Revised; Wechsler, 1974.

low-achievement definition. In the current sample, 19 of 32 RD-D children also met the low-achievement definition. The LA children, on the average, had lower IQ scores, but had reading scores comparable to the RD-D group (see Table 2).

Table 2 provides mean age and Woodcock–Johnson (1977) reading-cluster Rasch-scale scores at each grade for the three groups. Comparison of the mean reading-cluster Rasch-scale scores over the 9 years reveals that each group conforms to a pattern of diminishing gains with age, a pattern suggesting non-linear change at the group level given the absence of ceiling effects on the test. The differences in reading scores between the two disabled reading groups are clearly not statistically significant at any grade.

Procedure

To determine the most appropriate growth model for the data, we plot in Figure 1 the actual data for each child on the reading-cluster score against age at testing for all available assessments. As we hypothesized, this figure shows a pattern of more rapid change earlier in age, which slowed as children grew older. This pattern is curvilinear and implies that a quadratic model would be necessary to model growth. To evaluate the need for the quadratic parameter in the model, we performed a multilevel analysis (Francis et al., 1991) of reading-cluster Rasch-scale scores. The model included a linear and quadratic term for age to test the hypothesis that the data are fit by a curved line.

For these analyses and for all subsequent analyses, age was centered at age 8. This choice is arbitrary, but it was made because it reflects the mean age of children in Grade 2, a time before the identification of children as disabled, and is within the range of the data, which improves the estimation process by reducing the correlation among parameters. The practical implication of centering age at 8 years is that the intercept term in the model reflects expected reading performance at age 8, rather than at age 0 (i.e., birth). Similarly, in a quadratic model, the slope parameter repre-

Table 2
Means and Standard Deviations for Reading-Cluster Scores and Age at Testing for Three Reading Groups at Each Time Point

Variable	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
				Not reading impaired					
Reading									
M	462.0	486.5	499.6	508.6	515.0	521.8	527.3	531.5	535.8
SD	20.7	16.1	12.2	12.8	12.5	12.2	11.1	11.0	10.7
Age									
M	6.9	7.9	8.9	9.9	11.0	11.9	12.9	13.9	14.9
SD	0.3	0.3	0.3	0.4	0.4	0.4	0.3	0.3	0.4
n	333	331	333	334	334	332	329	329	328
				Reading disabled–discrepant					
Reading									
M	428.0	451.2	467.4	481.0	490.8	496.1	504.8	508.1	511.9
SD	24.3	18.6	15.8	18.5	19.1	17.8	19.0	19.8	20.6
Age									
M	7.0	8.0	9.0	10.0	11.1	12.0	13.0	14.0	15.0
SD	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
n	32	32	32	32	32	32	32	31	32
				Low achieving					
Reading									
M	426.0	454.9	469.2	483.1	491.4	497.4	503.8	508.6	513.2
SD	16.9	9.5	6.7	10.3	9.8	9.2	7.8	9.4	8.6
Age									
M	7.1	8.0	9.1	10.1	11.1	12.0	13.1	14.0	15.1
SD	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
n	37	37	37	37	37	37	37	37	35

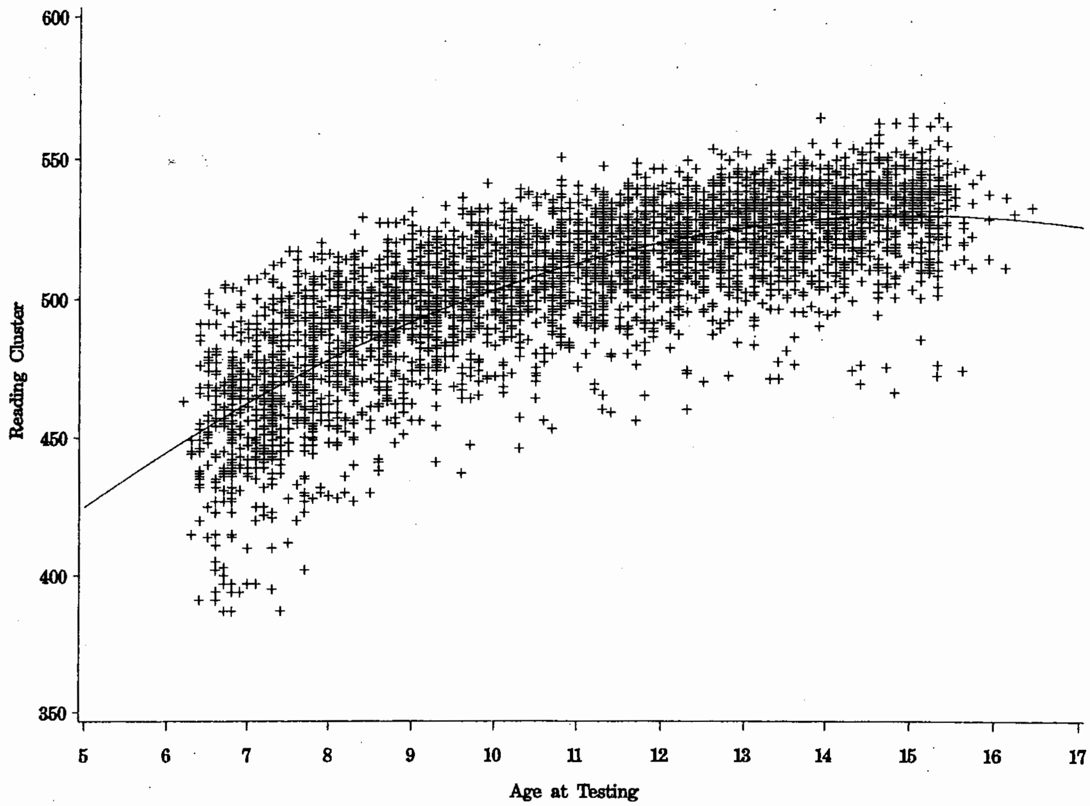


Figure 1. Age-related changes in the Woodcock–Johnson (1977) reading-cluster Rasch-scale score. Each case is plotted by age for all available assessments.

sents the instantaneous rate of change when the predictor is 0. Thus, by centering age at 8 years, the slope parameter represents the instantaneous rate of change at age 8, rather than at birth.

In the multilevel analysis, both the linear and quadratic terms were significant: linear, $t(3, 596) = 63.23, p < .0001$; quadratic, $t(3, 596) = 38.09, p < .0001$. The fact that the quadratic term was statistically significant indicates that there is curvature in the slope of the function representing change over time. In this case, the rate of change from year to year decreases as the children get older, which cannot be accounted for by a simple model that assumes a constant rate of change over time. The significance of the linear term indicates that there is some positive rate of improvement in reading levels at age 8. The mean slope was 14.9, and mean quadratic term was -1.07. The estimated variances of intercept, slope, and quadratic parameters were all statistically significant. Variance in these parameters indicates that the children differed in expected level of performance (intercept) at age 8, the rate of change at age 8 (slope), and the rate at which growth is decelerating (quadratic).

These analyses showed that in contrast to assumptions by Williamson et al. (1991), a linear model was not adequate to explain the relationship of reading and age. Rather, a model with a quadratic parameter is necessary. However, one potential problem with a simple quadratic model is that the underlying mathematical model is parabolic and predicts that performance will eventually turn back down and begin to decline. This mathematical form provides an unrealistic representation over the age range from grade school through high school and college. Rather, it would be expected that reading performance reaches a plateau at some age and then remains constant thereafter.

Figure 2 shows this distinction graphically. In Figure 2, it i clear that attempts to extrapolate beyond the data with the pur quadratic model would eventually be problematic because the pur quadratic model predicts that reading performance would eventu ally begin to decline. Modification of the quadratic model to allo growth to a plateau clearly provides a more realistic representatio of the growth process. Such a model would hypothesize th children grow at a steadily decreasing rate until a plateau achieved. Beyond the age at which the plateau is achieved, readin scores remain constant. These theoretical specifications sugge the following model:

$$y_{it} = \pi_{0i} + \pi_{1i}a_{it} + .5\pi_{2i}a_{it}^2 + R_{it} \text{ for age} < a_{ip} \qquad ($$

and

$$y_{it} = P_i + R_{it} \text{ for age} \geq a_{ip}, \qquad ($$

where y_{it} is reading performance for person i at time t, a_{it} is age f person i at time t, π_{0i} is expected reading performance for pers i when $a_{it} = 0$ (i.e., age 8), π_{1i} is the rate of change in readi performance for person i when $a_{it} = 0$, π_{2i} is the rate of decel ation in reading performance for person i, R_{it} is error in modeli performance for person i at time t, P_i is the level at which readi scores plateau for person i, and a_{ip} is the age at which sco plateau for person i.

The model of Equations 1 and 2 is inherently nonlinear beca of the constraint between the two equations. The constraint t Equations 1 and 2 are continuous at a_{ip} implies that

$$P_i = \pi_{0i} + \pi_{1i}a_{ip} + .5\pi_{2i}a_{ip}^2$$

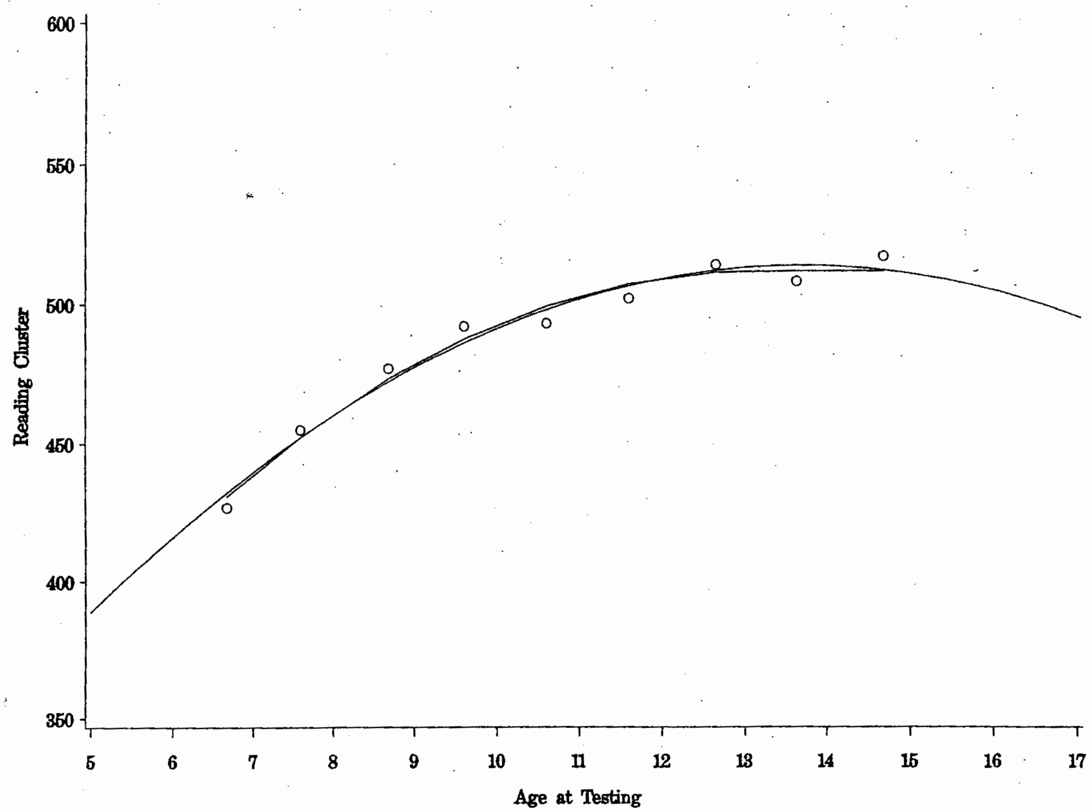


Figure 2. Quadratic models of change with and without plateau for a randomly selected child.

and

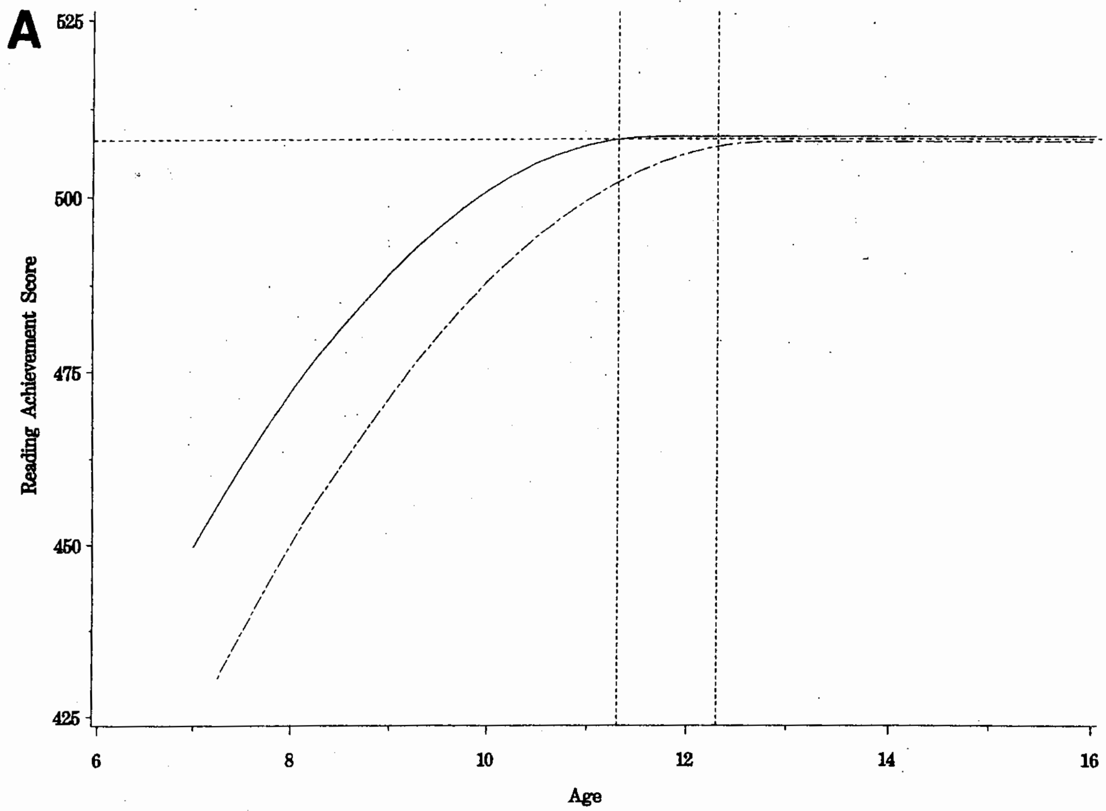
$$a_{ip} = -\pi_{1i}/\pi_{2i}, \quad (4)$$

where P_i and a_{ip} are as previously defined.

Our interest in fitting the model in Equations 1 and 2 was in estimating both the level at which performance plateaus (P_i), which is given by Equation 3, and the age at which the plateau is achieved (a_{ip}), which is given by Equation 4. To see how these parameters translate into hypotheses about lag and deficit, it is instructive to examine figures depicting hypothetical lag and deficit situations.

To help clarify what these equations mean for the lag–deficit hypothesis, we show in Figure 3 growth curves for two hypothetical groups. The horizontal reference lines depict the level at which performance plateaus; whereas the vertical reference lines mark the age at which the performance plateau is reached. The top of Figure 3 presents hypothetical growth curves for two groups in which the second group lags behind the first but eventually catches up. In this case, performance plateaus at the same level for both groups (note the equivalence of the horizontal reference lines), but the second group reaches its plateau at a later age (note the difference in the vertical reference lines). These groups would be found to differ in mean a_{ip} but not in mean P_{ip}. In contrast, consider the bottom of Figure 3. Here the second group lags behind the first but never catches up. The deficit in performance is consistent throughout the two growth curves and is ultimately reflected in the difference in levels at which performance plateaus (note the difference in horizontal reference lines), whereas there is no difference in the age at which the plateau occurs (note the single vertical reference line). These groups would be found to differ in mean P_{ip} but not in mean a_{ip}. Of course, it is possible that disabled readers lag behind nondisabled readers and never catch up. In this case we would expect the disabled readers to have greater mean a_{ip} and lesser mean P_{ip}. By fitting Equations 1 and 2 for every individual to estimate the parameters of Equations 3 and 4, the current study permitted evaluation of whether reading disabled and nondisabled readers differed in a manner consistent with a lag or deficit model and whether differences exist in this regard between disabled readers identified under two definitions of disability.

The models described in Equations 1 and 2 were fit separately for each individual to predict the Woodcock–Johnson (1977) reading-cluster Rasch-scale score with SAS PROC NLIN (Ihnen & Goodnight, 1989), by using the multivariate secant method of estimation. The model allowed individuals to differ in all growth parameters (intercept, rate of change, and rate of deceleration) and, consequently, the age and level at which achievement reached a plateau. Secondary analyses of the growth parameters were conducted by using SAS PROC GLM (Spector, Goodnight, Sall, & Sarle, 1989) ordinary least squares because most children had participated in an equal number of assessments with similar spacing. It would have been possible to improve precision of secondary analyses by differentially weighting the growth parameters based on the precision with which they had been estimated during the first analysis. Under common assumptions, these weights vary across children only on the basis of the differential number and spacing of assessments. The exact spacing of assessments was not identical across all cases because of fluctuations in the scheduling of annual assessments. Nevertheless, this difference was seen as relatively minor relative to the total variance in time points across the nine assessments. Moreover, the sample size was large, and the



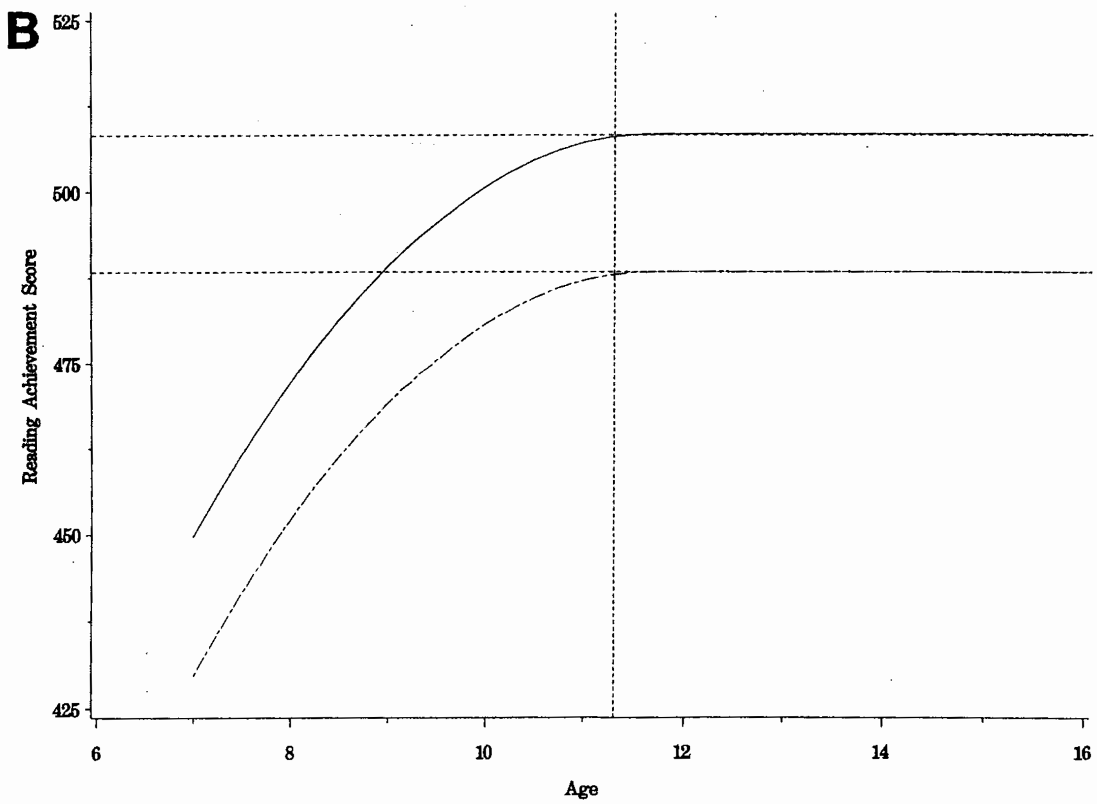


Figure 3. Hypothetical quadratic models of change with a plateau depicting a lag model (top) and a deficit model (bottom).

overall precision of the estimates indicated that power would be high for studying change correlates even without adjustment. In addition, 385 of 403 children were seen at all 9 time points. Of the remaining 18 children, 13 had 8 time points, 3 had 7 time points, and 2 had 6 time points. There was no relationship of these missing data to reading group assignment. For these reasons, we elected to conduct the second stage analysis with equal weighting for all children.

Results

Evaluating Fit of the Individual Growth Model

Before the specific hypotheses of interest are addressed, it is important to evaluate whether the model provides an adequate representation of the change process at the individual level. The first step in this evaluation is to examine the within-subject estimations, which revealed problems in estimation for 11 of 334 NRI, 1 of 32 RD-D, and 1 of 37 LA children. The problems included infinite standard errors or perfect correlation among one or more parameter estimates. Problems of this nature indicate that the model may be inappropriate for the case or that there are insufficient data to obtain good estimates. By adopting a weak model of cognitive development (Burchinal & Appelbaum, 1991), alternate model forms could have been fit to the data for these children. Because our interest in the current case was in evaluating the lag and deficit hypotheses, and such a small number of disabled children were affected, these 11 NRI and 2 disabled children were dropped from subsequent analyses.

Similarly, data from several cases returned growth parameter estimates that resulted in unrealistic achievement predictions. Unrealistic model predictions are another indication that the model may be inappropriate for a particular case. After estimation, cases were included in subsequent analyses only if the age at plateau was estimated to be between 5 and 23 years of age and the level of plateau was estimated to be between 200 and 800. This resulted in 2 LA, 3 RD-D, and 22 NRI cases being dropped from the analysis. Along with the cases lost because of missing or low IQ, this left a group of 34 LA, 29 RD-D, and 301 NRI children for testing the lag and deficit hypotheses. There was obvious concern that the dropping of cases could bias findings in favor of either the lag or the deficit hypothesis. This concern was greatest for cases dropped because the growth to a plateau model may be inappropriate. However, it should be pointed out that the total number of cases that were dropped for any reason was small (about 10%) and not related to group membership. Also, only very few cases were dropped because of the model being potentially inappropriate, and there appeared to be no relationship with group membership. Taken together, the small number of children involved and the lack of a relationship with group membership suggest that dropping these cases did not bias the test of the lag–deficit hypothesis. Consequently, it seemed reasonable to continue with a formal test of the lag and deficit hypotheses as operationalized in the quadratic growth to a plateau model.

Another aspect of model evaluation is to determine how well the model recovers the data for those cases where the model was deemed to be appropriate. To do so, squared multiple correlations were computed for each child individually by examining the difference between actual achievement and that predicted by the child's model at each age. Analogous to the squared multiple correlation in multiple regression, the sum of squared residuals was expressed as a percentage of the total corrected sums of squares in achievement and subtracted from 1.0. The squared multiple correlation estimates were excellent. In the LA group, R^2 ranged from .919 to .995; R^2 for the RD-D group ranged from .877 to .995; and for the NRI group, R^2 ranged from .845 to .997. Fewer than 4% of the cases in any group had R^2 values $< .90$.

Evaluating the Lag Versus Deficit Hypothesis

Descriptive statistics for the three growth parameters of Equation 4 are presented in Table 3 for each group. Table 3 indicates that the mean parameter values are quite similar for the RD-D and LA groups. To examine the statistical significance of the differences presented in Table 3 and to address the lag and deficit hypotheses, we conducted multivariate analyses of covariance (MANCOVAs) for the three growth parameters and separately for age at plateau and plateau level. For this second analysis, age at plateau and plateau level were first estimated for each individual by substituting the individual estimates of intercept, slope, and quadratic parameters into Equations 3 and 4. These estimates then served as the dependent variables for examination of the lag and deficit hypotheses.

To provide additional control over variation presumably not directly related to reading group, the three reading groups were also compared controlling for gender, race (White, African American, and other minority), and socioeconomic status (SES). Because of sample size limitations for some cells, only two levels of SES were considered (one and two vs. three, four, and five), and only main effects of factors were examined with the exception of the Gender × SES interaction. Because of missing data for SES and ethnicity, only 357 (of 363) cases were available for these adjusted analyses.

Although a number of factors affected both sets of outcomes, the presentation of results focuses on differences among the three reading groups. A MANCOVA of the three growth parameters indicated that reading groups were significantly different, $F(6, 680) = 24.15$, $p < .0001$. Inspection of the canonical structure matrix showed a large correlation only for the intercept parameter (.96) relative to the correlations for the slope (−.26) and quadratic (.10) parameters. Univariate significance tests suggest a similar conclusion, namely that the two disabled groups combined differed from the nondisabled group primarily in the level of reading ability across the age range. Reading groups differed on the intercept parameter (i.e., expected achievement at age 8), $F(2, 341) = 81.38$, $p < .0001$, with the NRI group showing higher reading levels. The slope parameter also

Table 3
Unadjusted and Least Squares Means and Standard Deviations of Growth Parameters for Three Reading Groups

Reading group	*n*	Intercept	Slope	Deceleration rate	Age at plateau	Plateau
				Unadjusted means		
Not reading impaired	301					
M		482.4	16.0	−2.9	15.0	533.7
SD		14.8	6.1	2.4	2.4	11.7
Reading disabled–discrepant	28					
M		450.6	18.3	−3.0	14.9	510.9
SD		17.7	5.2	1.7	1.7	21.9
Low achieving	34					
M		448.4	20.1	−3.5	15.0	512.3
SD		11.9	8.0	2.7	2.3	11.0
				Least squares means		
Not reading impaired	301	478.4	16.2	−2.8	14.9	530.5
Reading disabled–discrepant	28	449.9	18.5	−3.0	14.8	510.9
Low achieving	34	450.8	20.8	−3.7	14.9	516.2

Note. Reading groups differed in intercept and slope and in level of plateau. Disability groups did not differ from one another in the three growth parameters, level of plateau, or age at plateau. Results pertaining to reading group were unchanged by inclusion of gender, SES, and race. SES = socioeconomic status.

yielded a small but statistically significant effect, $F(2, 341) = 7.38$, $p < .0001$, with the combined LA and RD-D groups showing slightly higher slopes than the NRI group. There were no differences among groups in the quadratic effect, $F(2, 341) = 1.68$, $p < .19$.

The difference in slopes among the three reading groups is partly explained by the lower overall performance at age 8 of the two disabled groups relative to the nondisabled group. Because growth is curvilinear and decelerating, it is typical that lower performing individuals will show higher rates of change. This phenomenon is sometimes referred to as *the law of initial values* (Rogosa & Willett, 1985). In fact, when the three reading groups were compared on slopes controlling for intercepts and quadratic parameters, differences in slopes among the three groups were smaller ($p <$ .049), with the NRI group having the highest mean slope. Thus, when groups were equated for overall level of performance and the rate of deceleration, the nondisabled readers showed the highest instantaneous rate of growth at age 8. Although the slope differences among groups suggest that the disabled readers made up some of the differences in achievement levels that existed between them and the nondisabled readers, the comparison of slopes adjusted for level of performance and rate of deceleration suggests that their gains were smaller than those of nondisabled readers performing at comparable levels of the test.

Multivariate follow-up tests indicated that the two reading disability groups were not significantly different from each other ($F < 1$). Similar results were apparent for the univariate tests, intercept, $F < 1$; slope, $F(1, 47) = 1.15$, $p <$ .29; and quadratic parameter, $F < 1$. Effect size estimates for differences between disability groups in growth parameters were all small (Cohen's standardized effect size estimate, $f = .07$, $f = .13$, and $f = .11$, for intercept, slope, and quadratic parameters, respectively; Maxwell & Dela 1990).

The results for the growth parameters suggest that analysis of achievement plateau and age at plateau wil consistent with a model of developmental deficit in rea skill and inconsistent with a model of developmental la disabled children lag behind their nondisabled peers, differences should occur in the age at which the platea reached but not in the plateau itself. In contrast, the de model would predict a difference in plateau level but n the age at which the plateau is reached. The MANCOV age at plateau and plateau level also indicated a signifi effect of group, $F(4, 682) = 25.03$, $p < .0001$. Inspectic the canonical structure matrix showed a much higher relation for the plateau level (.86) relative to age at pla (.02). Univariate follow-up tests were also consistent this interpretation. The effect of age at plateau was significant ($F < 1$), whereas differences in plateau were significant, $F(2, 341) = 43.34$, $p < .0001$. Differe in age at plateau were very small (Cohen's standarc effect size $f = .02$), whereas differences in plateau were considerably larger ($f = .53$). Reading skills in NRI groups reached a plateau at a higher level thar combined RD-D and LA groups (see Table 3). When ing disability groups were compared separately, the n variate test was not significant ($F < 1$). In addition, ne of the univariate tests was significant: plateau level, 41) = 1.97, $p < .17$, and plateau age, $F < 1$. Differe between the two disability groups resulted in very estimated effect sizes ($f = .02$ for a_{ip}; $f = .04$ for P_i)

All analyses were repeated without controlling for SES, or gender, with no changes in the pattern of result illustrate these results, we used values for the unadj mean growth parameters from the first three colum

Table 3 to plot curves for each of the three reading groups. As Figure 4 shows, the major difference between disabled and nondisabled groups was the level at which the scores plateau. The slight difference in age at which the plateau occurs for the NRI versus the two reading disability groups was not statistically significant.

One possible problem with the comparison of the two reading-disabled groups arises from the inclusion of children in the discrepant group who were not low achieving (i.e., whose reading scores were above the 25th percentile). It can be argued that excluding these children from the discrepant group would provide a more direct comparison of discrepant and nondiscrepant disabled readers. To address this concern, we compared the RD-D and LA groups by using the separate variance *t* test, excluding the children in the RD-D group who scored above the 25th percentile in reading. A separate variance *t* test was used because of the difference in sample size in the two groups and the increased sensitivity of the pooled variance *t* to unequal variances when sample sizes are unequal. Moreover, the pattern of pairing sample variances and sample sizes would suggest that the pooled variance *t* was liberal in this particular situation (Maxwell & Delaney, 1990). (In fact, in such an analysis, the pooled variance test was $t(47) = -2.72$, $p = .009$.) The separate variance *t* test for plateau level was statistically significant at $p = .049$, with the RD-D group having the lower mean plateau. Differences for age at plateau were neither statistically ($p > .2$) nor clinically significant. The estimated difference in age at plateau was less than one half of a year. Thus, differences between groups were characterized by differences in the severity of the reading disability, which is consistent with the notion that disability exists on a continuum of reading development and that children who are IQ discrepant are the lowest achievers at any given level of IQ.

A possible second problem with these analyses concerns the nature of the outcome measure. The reading-cluster score of the Woodcock–Johnson (1977) is a composite reading measure composed of two measures of decoding and one measure of comprehension. It is possible that one

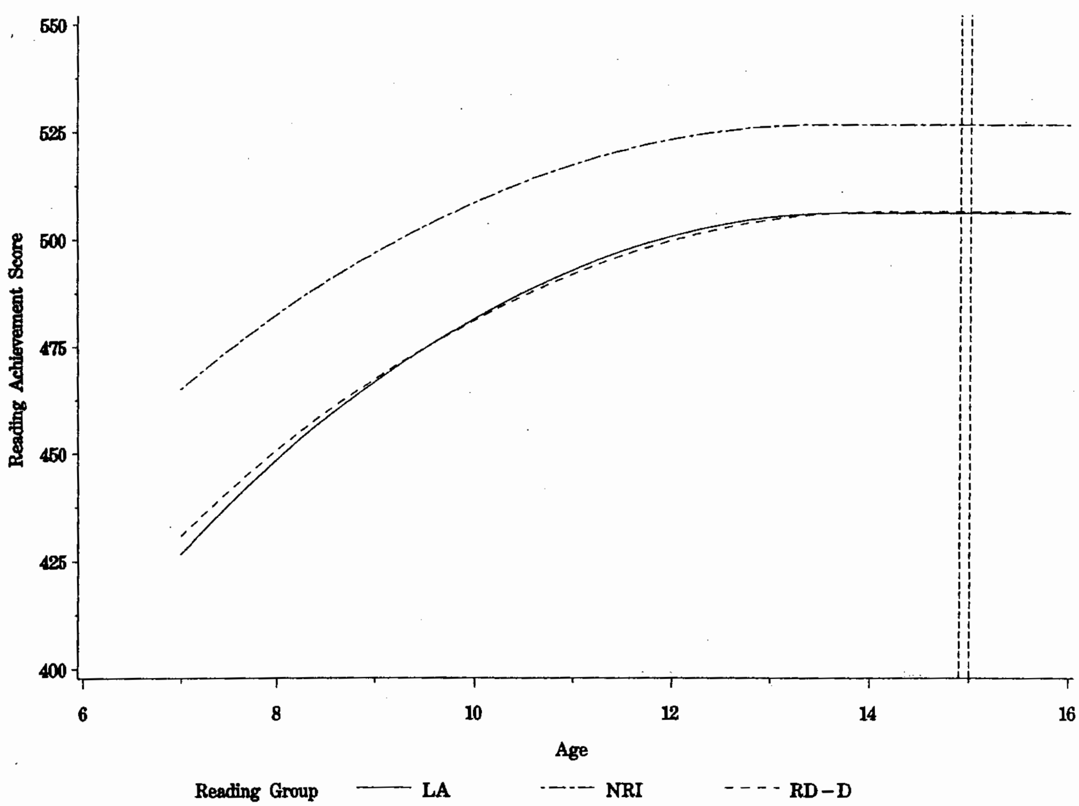


Figure 4. Estimated mean growth curves for the Woodcock–Johnson (1977) reading-cluster Rasch-scale score for the reading disabled–discrepancy (RD-D), reading disabled–low achievement (LA), and not reading impaired (NRI) groups. The curved lines show predicted achievement at each age by using mean growth parameters from the quadratic model with a plateau. The vertical reference lines are plotted at the estimated age at plateau for the NRI (left line) and disability (right line) groups. Reference lines for the LA and RD-D groups are coincident. The NRI group has a higher reading plateau but does not differ from the RD-D and LA groups in age at plateau. The disability groups do not differ from one another in reading plateau nor in age at plateau. These data are consistent with a deficit model of reading disability.

skill followed a lag model, whereas the other skill followed a deficit model, with one type of skill exerting greater influence on the analysis of the composite measure. To evaluate this possibility, we conducted separate analyses for each of the three Woodcock–Johnson subtest measures that make up the reading-cluster score. Results of these separate analyses resulted in identical conclusions to the analysis of the composite measure. In each case, differences were best characterized by deficits in reading skill for disabled readers, and the two disabled reader groups were not different from one another on any of the parameters of the model.

The analyses presented in this study have focused on changes in summary performance scores on the Woodcock–Johnson (1977) reading tests. Equally interesting are the responses that children made to specific items on the test and how these responses changed as children developed. To date, we have not conducted any detailed analyses of the specific errors or patterns of errors on the Woodcock–Johnson committed by the children in the CLS and have not examined how these errors might change over time. Although such an investigation would no doubt prove interesting, the children's original responses have not been coded electronically, which renders the task of error analysis somewhat formidable for the 9 years of CLS data that have currently been collected.

Discussion

This application of individual change models to the development of children with reading disabilities shows that models of growth to a plateau characterized the form of reading-skill development in children with reading disabilities and in children who are adequate readers. The models selected for evaluation were based on empirical examination of the individual and group growth curve data and assumptions about the nature of academic performance. Specifically, it was hypothesized that the rate of growth and development would vary across children and that children would vary in their final level of performance and the age at which they reached that level. Moreover, it was hypothesized that as children approach this ultimate individual performance level, their rate of development would continuously slow. Finally, it was hypothesized that performance would remain constant once this final level of performance had been achieved. Taken together, the assumptions lead to the formulation of a model that can be characterized as quadratic growth to a plateau. Within this formulation, the notion of developmental lag could be operationalized as a difference in the age at which two groups reach their final level of performance. In contrast, the notion of developmental deficit was operationalized as a difference in the final level of performance and may or may not be associated with any lag in development.

These hypotheses about the nature of change in reading skills over time were evaluated with individual growth curve analyses. The results clearly supported the hypothesis that the development of children with reading disabilities is best characterized by models involving a deficit. There was evidence for a small but statistically significant difference in rates of change at age 8 between disabled and nondisabled readers, with disabled readers having higher rates of change. This slope difference resulted in a reduction in the deficit for both disability groups from age 8 to the age at which reading scores reached a plateau. However, it is important to add that nondisabled readers showed higher rates of change at age 8 than the disabled readers when group differences in intercepts (i.e., level of performance at age 8) and rate of deceleration were controlled. Moreover, there were no differences among groups in the age at which scores reached a plateau, suggesting that groups did not differ, on average, in the duration of reading skill development.

In terms of the theoretical implications, the emergence of a deficit model does not support the hypothesis that reading disability results from a lag in brain maturation. The results also show that disabled readers fail to develop adequate reading skills, implying a problem that persists into adolescence and, in other studies, adulthood (Bruck, 1992). Intervention is essential, presumably at an earlier age in order to impact the persistence of poor reading. As we discuss below, individual growth models may be particularly useful for the early identification of children who are likely to develop reading disabilities.

There was also no evidence that different developmental models were necessary to model change in normal children or children who met discrepancy or low-achievement definitions of learning disability. Indeed, consistent with Fletcher et al. (1994), Stanovich and Siegel (1994), and with many other studies (e.g., Siegel, 1992), the two definitions of reading disability produced two groups that essentially did not differ. Moreover, discrepant low achievers differed from nondiscrepant low achievers only in intercept values and the level of the reading plateau (i.e., the deficit in the discrepant low achievers was more severe). These findings are consistent with the notion of reading disability representing extreme cases in an otherwise normal distribution of reading achievement and IQ (S. E. Shaywitz et al., 1992) adding to a growing body of literature that fails to support the hypothesized distinction between specific and garden-variety disabled readers when this distinction is based on IQ discrepancy. Given these results, children should be identified for services on the basis of either definition or an alternative approach that addresses the core deficit in phonological processing that clearly characterizes both groups (Fletcher et al., 1994; Stanovich & Siegel, 1994).

Groups were not found to differ in estimated age at plateau. For each of the three groups, reading scores were estimated to plateau on average at or around 15 years. This finding should not be taken as evidence that reading skills cease to develop by age 15. Indeed, there appeared to be considerable variability in the age at plateau. This variability is evidenced in Figure 5, which displays the level at which reading-cluster scores were estimated to plateau against the estimated plateau age of the child. Horizontal spread in the distribution of points indexes the heterogeneity in the estimated age at plateau. Clearly, a large group of plateaus was estimated to occur at age 15 or beyond. A

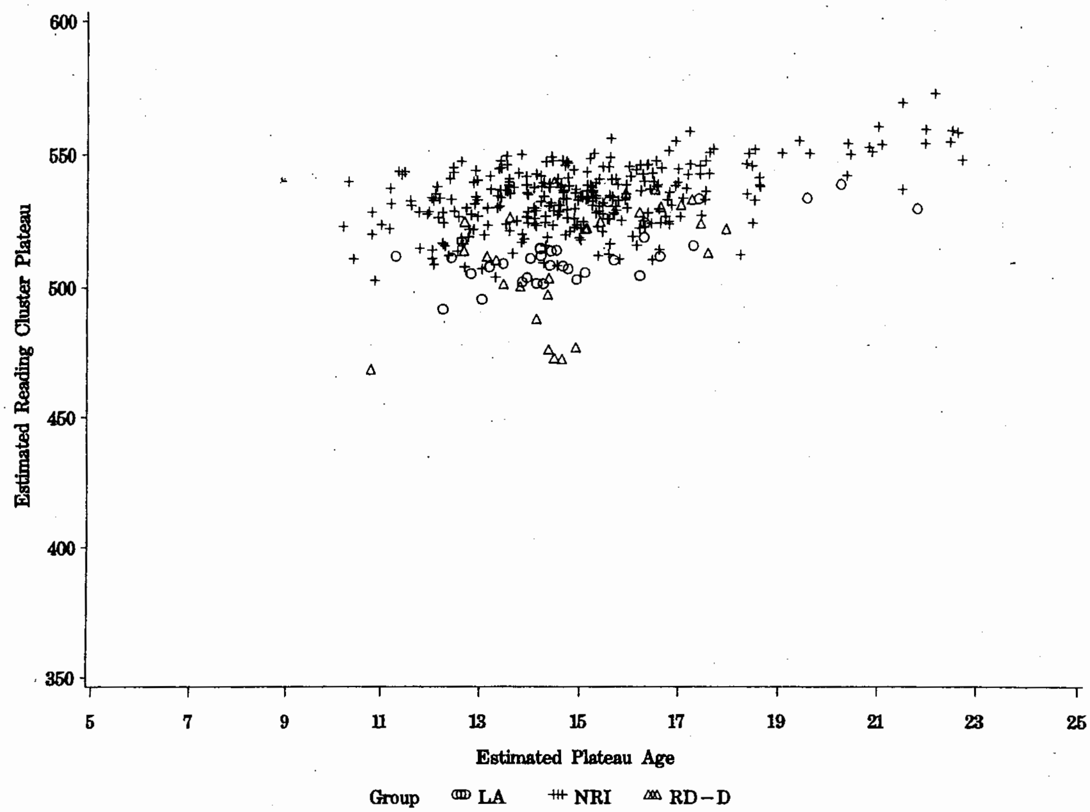


Figure 5. Relationship of estimated reading-cluster plateau and age at plateau by group. There is considerable variability in age at plateau for the low-achievement (LA), Reading disabled–discrepancy (RD-D), and not reading impaired (NRI) groups.

more data are collected on this cohort of children, evaluation of the accuracy of these predictions must be completed.

Finally, nothing in this research should be taken to suggest that reading skills cannot be developed beyond a certain age. In our study we used an individualized achievement measure to track development in reading in an unselected cohort of children. No attempt was made to evaluate the responsiveness of skills that have reached a plateau to intervention. Moreover, the fact that the fitted model does not directly estimate developmental parameters should caution against drawing firm conclusions about the continuance of the estimated trajectories beyond the observed time frame. Although the fit of the model to the available data was quite good, the empirical nature of the model selection process dictates that the predictions of the model be examined at further time points.

The capacity of individual change models to address issues concerning the development of reading skills in relationship to the definition of learning disabilities and assessment of outcomes was clearly apparent in this study. There is now widespread concern over current psychometric approaches to the definition and diagnosis of reading disability (Fletcher & Morris, 1986; Stanovich, 1991). At the me time, there is little consensus on how best to resolve ..e current problems in defining learning disabilities. The discussion of alternatives has focused on methods based strictly on cross-sectional assessments of ability development. An alternative is to incorporate longitudinal assessments into the diagnostic process. Given appropriate instruments, models, and analytic techniques, it is reasonable to propose to use individual growth parameters as the basis for validly and reliably identifying children who are not developing skills at rates sufficient to produce fluent reading by some specified age or grade. These models can also be used to assess outcomes in children as a function of intervention, ethnicity, curriculum, and a host of other factors.

The use of longitudinal assessments in the identification process and assessment of outcomes offers several compelling advantages over current static cross-sectional assessments. First, a focus on developmental functions and change emphasizes the measurement of learning over the measurement of learning outcomes. Such a shift places learning (as opposed to achievement) at the heart of defining learning problems and measuring outcomes, which is a conceptual advantage over current approaches because of the direct implications for intervention. In particular, intervention studies evaluating outcomes should incorporate methods for the assessment of individual change. Second, the use of longitudinal assessments to measure developmental rates makes early detection of learning problems a realistic possibility (Satz & Fletcher, 1988). Rather than wait until achievement outcomes can be measured, typically at the end of Grade 2, the development of precursor skills (e.g., phonemic segmentation) can be measured longitudinally, even in preschool, kindergarten, or Grade 1 years, so that children with poor acquisition rates can be identified. Moreover, rates of learning can be simultaneously measured in several skill domains to assess the degree to which skills are differentially developing. Such an assessment may help to distinguish failures to learn because of background and environmental conditions, such as poor motivation and poor parental attitudes about school, from those failures that are due to more discrete disabilities affecting specific cognitive domains such as deficits in phonological awareness.

Establishment of such a growth rate model for identification and assessment of outcome requires instruments suitable for interval scaling of individuals on the relevant constructs. Without interval or ratio-scaled data, children showing the lowest growth rate may not represent those actually growing at the lowest rate. The need for interval scale assessments presents a major stumbling block for the implementation of individual change models. Psychometrically sound instruments exist for cognitive assessment. Unfortunately, instrument developers have not focused on the measurement of change and may delete those items showing the most change over time because of the preoccupation of normative comparisons of same-aged children. This limits the capacity of the instrument to detect change. Studies of the phenomenon of change should include careful examinations of the device used to indicate change to ensure that it can appropriately be scaled. Although some simple methods exist for utilizing available instruments in the study of change, many measures of academic and cognitive skills are not suitable because the scoring metrics will not yield values on an interval or ratio scale. Application of item response models to existing instruments to develop alternate scoring methods may allow for cost effective development of test instruments suitable for studying change. A strength of the present study is the Rasch-scale scoring available for the Woodcock–Johnson (1977), which ensures that performance is measured on an interval scale.

It is possible that experimental designs in which older disabled readers are compared to younger nondisabled readers who are matched on reading level would show results more compatible with a developmental lag model. However, such a conclusion was not supported by Stanovich and Siegel (1994). In this longitudinal study, we extended their findings to a situation in which change was *nonlinear,* a situation that threatens the external validity of the analytic methods and cross-sectional design used by Stanovich and Siegel. Using reading-level match designs to make inferences and test hypotheses about change remains an inherently limited approach to the understanding of development.

It is also possible that results supporting a lag model would emerge if analyses of change were conducted for outcome measures other than reading. This seems unlikely for measures with strong relationships to reading ability and disability (e.g., phonological awareness). Although the analyses reported here focused on a composite measure of reading, similar findings were obtained when the individual components of that measure were analyzed separately. Studies of adults with a history of reading disability have also failed to show resolution of either the word decoding or the phonological processing problems associated with reading disability (Bruck, 1992). In the present study, models involving growth to a plateau fit children at all levels of ability, with only small and nonsignificant differences in age at plateau between ability groups. Although the model of growth to a plateau characterized growth quite well for children in all reading groups at all levels of ability, there was a small number of children in each group (less than 10%) for whom the model may not have been appropriate. These children certainly warrant further in depth study and continued follow-up. Also of particular interest is the striking variability in age at plateau across the children in this study. Although this variability was not related to reading group, further examination of possible factors influencing when the development of reading ability plateaus would seem fruitful for future investigations.

References

Backman, J., Bruck, M., Hebert, M., & Seidenberg, M. (1984 Acquisition and use of spelling sound correspondences in read ing. *Journal of Experimental Child Psychology, 38,* 114–133

Backman, J., Mamen, M., & Ferguson, H. (1984). Reading lev design: Conceptual and methodological issues in reading r search. *Psychological Bulletin, 96,* 560–568.

Beech, J., & Harding, L. (1984). Phonemic processing and t poor reader from a developmental lag viewpoint. *Reading R search Quarterly, 19,* 357–366.

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

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hiera chical linear models to assessing change. *Psychological Bullet 101,* 147–158.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individu developmental functions: Methods and their assumptions. *Ch Development, 62,* 23–43.

Cromer, W. (1970). The difference model: A new explanation some reading difficulties. *Journal of Educational Psycholo 61,* 471–483.

Fletcher, J. M. (1981). Linguistic factors in reading acquisition. F. Pirozzolo & M. Wittrock (Eds.), *Neuropsychology and co nitive processes in reading* (pp. 261–294). New York: Acader Press.

Fletcher, J. M., & Morris, R. (1986). Classification of disab readers: Beyond exclusionary definitions. In S. Ceci (E *Handbook of cognitive, social, and neuropsychological aspe of learning disabilities* (Vol. 1, pp. 55–80). Hillsdale, Erlbaum.

Fletcher, J. M., Satz, P., & Morris, R. (1984). The Florida Lor tudinal Project: A review. In S. A. Mednick, M. Harway, K. M. Finnello (Eds.), *Handbook of longitudinal research* (1, pp. 280–304). New York: Praeger.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, Liberman, I. Y., Fowler, A., Francis, D. J., Stuebing, K. K. Shaywitz, B. A. (1994). Cognitive profiles of reading disabi Comparisons of discrepancy and low achievement definiti *Journal of Educational Psychology, 85,* 1–18.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Davidson, K. C Thompson, N. M. (1991). Analysis of change: Modeling i vidual growth. *Journal of Consulting and Clinical Psychol 59,* 27–37.

Francis, D. J., Shaywitz, S. A., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1994). The measurement of change: Assessing behavior over time and within a developmental context. In G. R. Lyon (Ed.), *Frames of reference: Measurement of learning disabilities.* Baltimore, MD: Paul H. Brooks.

Frankenberger, W., & Fronzaglio, K. (1991). A review of states' criteria for identifying children with learning disabilities. *Journal of Learning Disabilities, 24,* 495–500.

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

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology, 65,* 294–299.

Guthrie, J. T., & Seifert, M. (1977). Letter-sound complexity in learning to identify words. *Journal of Educational Psychology, 69,* 686–696.

Guthrie, J. T., & Tyler, S. J. (1975). Psycholinguistic processing in reading and listening among good and poor readers. *Journal of Reading Behavior, 8,* 415–426.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Norwell, MA: Kluwer Academic.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Ihnen, L. A., & Goodnight, J. H. (1989). The NLIN procedure. In *SAS/STAT user's guide* (Vol. 2, Version 6, 4th ed., pp. 1135–1193). Cary, NC: SAS Institute.

Ishwaran, H., Shaywitz, B. A., Francis, D. J., Escobar, M. D., Fletcher, J. M., Holahan, J., & Shaywitz, S. E. (1995). *IQ/Achievement discrepancy scores in grades 1–8: A structural equations approach to understanding reliability.* Manuscript submitted for publication.

Jackson, N. E., & Butterfield, E. C. (1989). Reading-level-match designs: Myths and realities. *Journal of Reading Behavior, 21,* 387–412.

Maxwell, S. E., & Delaney, H. A. (1990). *Designing experiments and analyzing data: A model comparison perspective.* Belmont, CA: Wadsworth.

Olson, R., Kliegl, R., Davidson, B., & Foltz, G. (1985). Individual and developmental differences in reading disability. In T. Waller (Ed.), *Reading research: Advances in theory and practice* (Vol. 4, pp. 1–64). London: Academic Press.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 90,* 726–748.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rourke, B. P. (1976). Reading retardation in children: Developmental lag or deficit. In R. Knights & D. J. Bakker (Eds.), *The neuropsychology of learning disorders: Theoretical approaches.* (pp. 125–137). Baltimore, MD: University Park Press.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16,* 181–197.

Satz, P., & Fletcher, J. M. (1980). Minimal brain dysfunctions: An appraisal of research concepts and methods. In H. Rie & E. Rie (Eds.), *Handbook of minimal brain dysfunctions* (pp. 669–714). New York: Wiley.

Satz, P., & Fletcher, J. M. (1988). Early identification of learning disabilities: An old problem revisited. *Journal of Consulting and Clinical Psychology, 56,* 824–829.

Satz, P., Fletcher, J. M., Clarke, W., & Morris, R. (1982). Lag, deficit, delay, and rate constructs in learning disability: A re-examination. In A. Ansara, N. Geschwind, A. Galaburda, M. Albert, & N. Gartrell (Eds.), *Sex differences in dyslexia* (pp. 129–150). New York: Orton Dyslexia Society.

Satz, P., Taylor, H. G., Friel, J., & Fletcher, J. M. (1978). Some developmental and predictive precursors of reading disabilities: A six year follow-up. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge* (pp. 313–347). New York: Oxford University Press.

Share, D. L., & Silva, P. A. (1986). The stability and classification of specific reading retardation: A longitudinal study from age 7 to 11. *British Journal of Educational Psychology, 56,* 32–39.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30,* 894–906.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326,* 145–150.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association, 264,* 998–1002.

Siegel, L. S. (1992). Dyslexics vs. poor readers: Is there a difference? *Journal of Learning Disabilities, 25,* 618–629.

Spector, P. C., Goodnight, J. H., Sall, J. P., & Sarle, W. S. (1989). The GLM procedure. In *SAS/STAT user's guide* (Vol. 2, Version 6, 4th ed., pp. 891–996). Cary, NC: SAS Institute.

Stanovich, K. E. (1988). Explaining the difference between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21,* 590–612.

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

Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly, 21,* 267–283.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The developmental lag hypothesis in reading: Longitudinal and matched reading-level comparisons. *Child Development, 59,* 71–86.

Stanovich, K. E., & 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, 86,* 24–53.

Vellutino, F. R., & Scanlon, D. M. (1989). Some prerequisites for interpreting results from reading level matched designs. *Journal of Reading Behavior, 21,* 361–385.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised.* San Antonio, TX: Psychological Corporation.

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analysis of academic achievement. *Journal of Educational Measurement, 28,* 61–76.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock–Johnson Psychoeducational Test Battery.* Boston: Teaching Resources.

Received September 8, 1994
Revision received October 10, 1995
Accepted October 11, 1995 ■