Journal of Educational Psychology
1996, Vol. 88, No. 4, 601–638

0022-0663/96/$3.00

Cognitive Profiles of Difficult-to-Remediate and Readily Remediated Poor Readers: Early Intervention as a Vehicle for Distinguishing Between Cognitive and Experiential Deficits as Basic Causes of Specific Reading Disability

Frank R. Vellutino, Donna M. Scanlon, Edward R. Sipay, and Sheila G. Small
University at Albany, State University of New York

Alice Pratt
Niskayuna Central School District, Niskayuna, New York

RuSan Chen
University at Albany, State University of New York

Martha B. Denckla
Johns Hopkins University

Reading impaired first graders were given daily tutoring as a "first cut" diagnostic to aid in distinguishing between reading difficulties caused by basic cognitive deficits and those caused by experiential deficits. Reading achievement in most of these children was found to be within or above the average range after one semester of remediation. Children who were difficult to remediate performed below both children who were readily remediated and normal readers on kindergarten and first-grade tests evaluating phonological skills, but not on tests evaluating visual, semantic and syntactic skills. The results are consistent with convergent findings from previous research suggesting that reading problems in some poor readers may be caused primarily by phonological deficits.

Specific reading disability is conventionally defined as severe difficulty in learning to identify printed letters and words in children who have at least average intelligence and who are not impaired by general learning difficulties (Gough & Tunmer, 1986; Stanovich, 1988; Vellutino, 1979, 1987). As an etiological concept, it carries with it the implicit assumption that the reading problems of such children are caused primarily by constitutional factors such as organic disorder or genetic limitations that adversely affect cognitive abilities that underlie reading ability. Commonly used definitions of specific reading disability typically employ a number of exclusionary criteria defining experientially limiting factors such as low general intelligence, sensory deficits, emotional disorder, motivational problems, frequent absences from school, or socioeconomic impoverishment to aid in distinguishing between children whose reading difficulties are caused primarily by inadequate experience and children whose reading difficulties are caused primarily by basic deficits in cognitive abilities that underlie reading ability. However, aside from the fact that they provide no definitive means for distinguishing between these two groups, the use of such exclusionary criteria as the sole vehicle for identifying reading disability encounters at least two other problems. The first is that this approach does not necessarily screen out children whose reading difficulties might be caused primarily by inadequate schooling or limited exposure to reading readiness activities. This point is well articulated in a penetrating article by Clay (1987), who argued forcefully that the failure to control for the child's educational history is the major impediment to differential diagnosis of reading disability. Indeed, she suggested that virtually all studies that have sought to evaluate basic process deficit explanations of reading disability are confounded by this problem and aptly pointed out that the adverse effects of inadequate prereading experience, inadequate instruction, or both can often mask or even mimic the adverse effects of constitutionally based cognitive deficits.

Frank R. Vellutino, Donna M. Scanlon, Edward R. Sipay, Sheila G. Small, and RuSan Chen, Child Research and Study Center, University of Albany, State University of New York; Alice Pratt, Niskayuna Central School District; Martha B. Denckla, Kennedy Krieger Institute, Johns Hopkins University.

The data for this study were collected as part of a project conducted under the auspices of special Center Grant P50HD25806 to the Kennedy Krieger Institute of Johns Hopkins University from the National Institute of Child Health and Human Development. Martha Bridge Denckla was the principal investigator overseeing the various research projects initiated under this Center grant. The research reported here is part of Project IV (The Reading and Language Project) under a subcontract directed by Frank R. Vellutino and Donna M. Scanlon of the Child Research and Study Center of the University at Albany.

We would like to express our sincere appreciation to the teachers, students, administrators, and secretarial staff in the schools that participated in this study. Thanks are also due to the many members of the research staff at the Child Research and Study Center who collected, collated, and processed the data. Special thanks go to Diane Fanuele, Melinda Taylor, Judy Moran, Melissa Hollenberg, Debbie Tuohy, Margaret Feldman, and Nancy Eson.

Correspondence concerning this article should be addressed to Frank R. Vellutino, Child Research and Study Center, University at Albany, State University of New York, Albany, New York 12222. Electronic mail may be sent via Internet to FRV89@cnsibm.albany.edu.

This occurs partly as a function of knowledge gaps and ineffective learning strategies and partly as a function of encumbering social and emotional problems that often accrue in reading-impaired children.

Clay's (1987) concerns are given a good deal of credibility by intervention studies that have shown that most impaired readers can acquire at least grade-level reading skills if they receive early and labor-intensive intervention to correct their reading deficiencies (Clay, 1985; Iversen & Tunmer, 1993; Pinnell, 1989; Wasik & Slavin, 1993). Her concerns are also inherent in the second problem associated with the use of exclusionary criteria as the sole vehicle for distinguishing between constitutionally and experientially based causes of reading difficulties. The problem is that the constitutional–experiential dichotomy may, itself, lack ecological validity, at least in terms of the common stereotype that impaired readers, selected on the basis of exclusionary criteria, suffer from organic deficits that adversely affect one or another of the cognitive abilities that underlie reading ability. This may or may not be true in any given case, and the question of whether or not it may be true in some instances can only be addressed through empirical research. However, this and like stereotypes fail to reflect the fact that any given level of reading achievement is a by-product of a complex interaction between one's endowment and the quality of one's literacy experience and instruction, such that the child who is endowed with an adequate mix of the cognitive abilities underlying reading ability is better equipped to profit from experience and instruction in learning to read than is the child who is endowed with a less than adequate mix of these abilities. Indeed, the optimally endowed child may be able to profit from less than optimal experience and instruction, whereas the inadequately endowed child may have difficulty profiting from even optimal experience and instruction.

Through this analysis, the constitutional–experiential dichotomy may be usefully reconceptualized insofar as reading ability and the cognitive abilities underlying reading ability may be conceived of as continuous variables, whereas constitutionally based reading impairment may be conceived of as achievement at the low end of the reading ability continuum in children having reading-related cognitive abilities that are also at the low end of their (respective) continuua. Conversely, experientially based reading impairment may be conceived of as achievement at the low end of the reading ability continuum in children having reading-related cognitive abilities toward the middle or even upper end of their respective continuua, but who, nevertheless, have difficulty learning to read because of inadequate experience and instruction. Accordingly, the goal of reading disability researchers would seem to be threefold: first, to isolate cognitive abilities that are especially important for learning to read, along with deficiencies in these abilities that might distinguish between poor and normally developing readers; second, to isolate experiential and instructional variables that differentially affect achievement in reading; and third, to isolate the genetic and neurological underpinnings of the cognitive abilities underlying reading ability in the interest of distinguishing between genetic and neuropathological causes of deficits in these abilities.

Over the past few decades, researchers operating within this framework have made some progress in all three areas and, thus, in distinguishing probable from improbable causes of reading disability. To illustrate, specific reading disability has been attributed to dysfunction in selective attention (Douglas, 1972), associative learning (Brewer, 1967; Fildes, 1921; Gascon & Goodglass, 1970), cross-modal transfer (Birch, 1962), serial-order processing (Bakker, 1972), and rule learning (Morrison & Manis, 1982), but dysfunction in one or another of these rather basic and general learning abilities would seem to be ruled out as significant causes of the disorder in a child who has at least average intelligence and who does not have general learning difficulties, given that all of these cognitive abilities are entailed in virtually all tests of intelligence and are most certainly entailed in all academic learning. Moreover, the empirical evidence for etiological theories implicating one or another of them is weak (Vellutino, 1979, 1987; Vellutino & Scanlon, 1982). Reading disability has also been attributed to dysfunction in visual processing, as well as to deficiencies in the phonological, semantic, and syntactic domains of language. Although the evidence supporting phonological deficit explanations of reading disability is very strong and highly convergent, the evidence against most visual deficit explanations is equally strong and is also highly convergent. At the same time, the evidence for semantic and syntactic deficit explanations is mixed.

In brief, the results of a large number of studies, when taken together, permit the inference that reading disability, in many cases, is caused by phonological coding deficits that impair the acquisition of phonological skills such as phoneme segmentation, letter and word naming, letter-sound mapping, name retrieval, and verbal memory. This inference is supported by cross-sectional studies with both children and adults, in which poor readers were generally found to be less proficient than normal readers on measures evaluating these skills (Brady, Shankweiler, & Mann, 1983; Bruck, 1990, 1992; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979; Tunmer, 1989; Vellutino & Scanlon, 1982, 1987a, 1987b; Vellutino, Scanlon, & Spearing, 1995), as well as by longitudinal studies demonstrating that they are reasonably good predictors of achievement in reading (Adams, 1990; Blachman, 1984; Bradley & Bryant, 1983; Liberman, Shankweiler, Fischer, & Carter, 1974; Lundberg, Olofsson, & Wall, 1980; Vellutino & Scanlon, 1987b; Wolf, 1984; Yopp, 1995). Also supportive are regression studies demonstrating that tasks evaluating phonological skills account for more variance on measures of word identification than do tasks evaluating other language-based skills or those evaluating visual-processing abilities (Vellutino, Scanlon, Small, & Tanzman, 1991; Vellutino, Scanlon, & Tanzman, 1994). However, the most direct support for a causal relationship between phonological skills and reading ability comes from training studies demonstrating that direct instruction in phoneme segmentation and letter-sound mapping can improve word identification and spelling ability

(Ball & Blachman, 1991; Blachman, Ball, Black, & Tangel, 1994; Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1990, 1991; Foorman, Francis, Novy, & Liberman, 1991; Fox & Routh, 1980; Lundberg, Frost, & Petersen, 1988; Vellutino & Scanlon, 1987b; Williams, 1980).

Some support for a semantic deficit theory of reading disability is provided by studies in which impaired readers have been found to be less proficient than normal readers on tests of vocabulary development and semantic concept development (Bryan, Donahue, & Pearl, 1981; Donahue, 1986; Fry, Johnson, & Muehl, 1970; Kavale, 1982; Loban, 1963). However, the results of most of these studies are compromised by the fact that the participants evaluated either came from impoverished backgrounds or were generally impaired academically and not just in learning to read. Moreover, in studies more recently conducted by Vellutino and his associates (Vellutino & Scanlon, 1987a; Vellutino, Scanlon, & Tanzman, 1988; Vellutino et al., 1991; Vellutino et al., 1994; Vellutino et al., 1995), strong and reliable reader group differences on semantic measures were observed only in contrasts of poor and normal readers in sixth and seventh grade, but not in contrasts of poor and normal readers in second and third grade. This pattern of results is contrary to a semantic deficit theory of reading disability. It is more in keeping with Stanovich's (1986) suggestion that semantic deficits that may be observed in poor readers of the type typically studied in reading disability research—that is, otherwise normal children who are not generally impaired in learning—are among the many cognitive deficits that accrue as a consequence of reading disability (what Stanovich called "Matthew effects") and are not a primary cause of the disorder.

As regards syntactic deficit theories of reading disability, the research has been less conclusive. Poor readers have been found to be deficient, relative to normal readers, on tests evaluating (a) knowledge of inflectional morphemes (Brittain, 1970; Vellutino & Scanlon, 1987a; Vogel, 1974); (b) comprehension of complex syntax (Byrne, 1981; Goldman, 1976; Vellutino & Scanlon, 1987a; Vogel, 1974); (c) the ability to detect or repair grammatically ill-formed sentences, sometimes called "syntactic awareness" (Flood & Menyuk, 1983; Fowler, 1988; Tunmer, Nesdale, & Wright, 1987; Vellutino & Scanlon, 1987a); and (d) the ability to use sentence contexts to facilitate and monitor word identification (Guthrie, 1973; Tunmer & Hoover, 1993). However, the origin of such deficiencies is at issue. Thus, Mann, Shankweiler, and Smith (1984) and others (Shankweiler, Crain, Brady, & Macaruso, 1992) suggested that poor readers may be found to have difficulty on syntactic tasks such as comprehending complex sentences, judging grammaticality, or making use of sentence context for word identification because such tasks make heavy demands on working memory, and poor readers, Mann et al. asserted, have limited working memory capacity as a consequence of phonological coding deficits. Moreover, in a series of studies conducted by Mann et al., poor and normal readers who were distinguished on working memory and other phonological tasks were found to perform at comparable levels on syntactic tasks that did not tax working memory, for example, sentences that were less complex and contained fewer idea units (see Shankweiler et al., 1992, for a summary of this research).

These results provide tentative support for the view that syntactic deficits in many poor readers may be a consequence of deficiencies in phonological coding that abnormally limit working memory capacity. However, we should point out that syntactic deficits in some poor readers could also be a consequence of prolonged reading difficulties. In fact, Vellutino and Scanlon (1987a) found no statistically significant differences between second-grade poor and normal readers on measures of sentence comprehension after controlling for working memory differences in the two groups. Controlling for working memory did not, however, eliminate differences between sixth-grade reader groups on these measures, in accord with the possibility that syntactic deficits may accrue as a consequence of prolonged reading difficulties. Of course, these two possibilities are not mutually exclusive and the issue remains open.

Finally, reading disability has been variously attributed to dysfunction in visual memory, visual form perception, spatial orientation, and directional sequencing, in addition to inherent spatial confusion and visual tracking problems associated with oculomotor deficiencies (Getman, 1985; Hermann, 1959; Orton, 1925; Pavlidis, 1981). However, there is now abundant evidence that poor and normal readers tend not to differ on measures evaluating visuospatial abilities of the types just mentioned (Vellutino, 1979, 1987). Similarly, well-controlled studies evaluating visual tracking of nonverbal stimuli obtained no differences between these two groups, contrary to the notion that oculomotor defects cause reading disability (Olson, Kliegl, & Davidson, 1983; Stanley, Smith, & Howell, 1983).[1]

As regards constitutionally based causes of reading disability, results have been seminal but suggestive. Support for the possibility that reading disability may be related to inadequate endowment has come from genetic studies that have documented (a) that reading difficulties occur more often in near relatives than in the population at large, (b) that

[1] A number of investigators have also theorized that reading disability may be caused by a deficit in the "transient visual system," which is believed to be responsible for inhibiting the visual trace during saccadic movements of the eyes (Breitmeyer, 1989; Lovegrove, Martin, & Slaghuis, 1986; Martin & Lovegrove, 1984). This deficit is said to cause visual trace persistence in moving from one word to the next when reading connected text, thereby creating visual masking effects that impair reading of the text. However, as pointed out by Hulme (1988), the trace persistence explanation of reading disability predicts that poor readers should be impaired only when reading connected text and not when words are encountered one at a time under foveal vision (sustained system) conditions. Yet, it is known that poor readers have as much or more difficulty identifying printed words encountered one at a time under foveal vision conditions as they have identifying words encountered in connected text. Moreover, there is no evidence that poor readers are impaired by visual masking and visual acuity problems under normal reading conditions. Thus, evidence of these effects observed under laboratory conditions are, at most, epiphenomenal.

they occur more often in twins than in siblings, and (c) that they have a much higher concordance rate in monozygotic twins than in dizygotic twins (Decker & Vandenberg, 1985; DeFries, 1985; DeFries, Fulker, & LaBuda, 1987; Olson, Wise, Conners, Rack, & Fulker, 1989; Olson, Wise, Conners, & Rack, 1990). Moreover, a recent study has tentatively located a gene for reading disability on Chromosome 6, although this finding has not yet been replicated (Cardon et al., 1994). Finally, genetic and twin studies have shown that measures of reading ability as well as measures of phonological skills such as phoneme segmentation, letter-sound mapping, and rapid naming have high degrees of heritability (Olson et al., 1989; Olson et al., 1990). These findings have provided the most compelling evidence for a direct link between genetic endowment and reading-related cognitive abilities.

Neuroanatomical studies reported by Galaburda and Kemper (1979; see also Galaburda, 1983) have provided suggestive evidence that reading difficulties, in some cases, may be caused by neurodevelopmental anomalies. In postmortem analyses of the brains of adult men with a history of reading difficulties, these investigators found defects in the architecture of the language areas of the left hemisphere. They also found that the left hemisphere was no better developed than the right hemisphere, which is atypical. The possibility that left hemisphere anomalies may be causally related to reading disability is given added substance by electrophysiological studies conducted by Duffy, Denckla, Bartels, and Sandini (1980) and by Shucard, Cummins, Gay, Lairsmith, and Welanko (1985), who found that left hemisphere brain wave responses of dyslexic and normal readers engaged in various cognitive tasks were qualitatively different (see also Dykman, Ackerman, & Holcomb, 1985; Harter, Anllo-Vento, Wood, & Schroeder, 1988).

Finally, a small number of studies using both computed tomographic and magnetic resonance imaging procedures to compare the structural integrity of the brains of impaired and normal readers (see Hynd & Semrud-Clikeman, 1989, and Filipek, 1995, for reviews) have, in several instances, revealed neuroanatomical differences between these two groups. The nature and location of such differences vary from study to study, but are most frequently evident in the planum temporale, insular cortex, and corpus callosum of the brains of the impaired readers. However, results, to date, have been inconclusive.

To summarize to this point, research evaluating etiological theories of reading disability has provided much greater support for linguistic deficit than for visual deficit theories of reading disability, but strong and convergent evidence for a causal relationship has been garnered only for those theories implicating deficits in phonological skills as basic causes of the disorder. However, if Clay (1987) is correct in suggesting that virtually all research evaluating the etiology of reading disability is confounded by the failure to control for the child's educational history, then it must be acknowledged that this and other explanations of the disorder need to be reexamined after some attempt is made to effect such control. What remains to be done to address Clay's concerns and to validate previous findings is to compare selected samples of impaired readers on reading and reading-related cognitive tasks, at both the preliterate and beginning stages of reading development and both before and after intensive intervention. If reading disability in otherwise normal children is caused by basic deficits in skills that depend heavily on phonological coding ability, then the types of phonological tasks that have been found to reliably and robustly distinguish between poor and normal readers should also distinguish between poor readers who are difficult to remediate and poor readers who are readily remediated. In contrast, such tasks should less reliably and less robustly distinguish between normal readers and readily remediated poor readers. Conversely, visual tasks should not readily distinguish between difficult-to-remediate and readily remediated poor readers because such tasks have not reliably distinguished between poor and normal readers in previous research.[2] Finally, if semantic and syntactic deficits that have been observed in poor readers are consequences of reading disability rather than primary causes of the disorder, then there should be no appreciable differences between respective participant groups on tasks evaluating these abilities, either at the preliterate stage or at the beginning stage of reading development.

To test these hypotheses, we conducted a longitudinal study that continuously evaluated a large group of elementary school children from kindergarten through fourth grade. All kindergartners from participating school districts were administered a large battery of psychological tests evaluating relevant cognitive abilities and rudimentary literacy skills. In mid-first grade, subsamples of poor and normal readers were selected from the larger sample of children initially tested in kindergarten. These were our target children, and they were studied in-depth. Poor readers from this group were randomly assigned to either a tutored group or a nontutored "contrast" group designed to help evaluate the relative effectiveness of our daily tutoring program. Children in the tutored group were given daily tutoring (30 min per day) for between one and two school semesters, depending on progress. Children in the nontutored group were given school-based remediation. All target children were given individually administered achievement tests at least annually through fourth grade. They were also evaluated in first and third grade on a battery of tests assessing cognitive abilities believed to underlie reading ability. For purposes of comparison and cross-validation, the target children and children from the larger sample initially tested in kindergarten (and not lost through attrition) were also evaluated annually through group-administered school-based testing. However, in this article, we report results primarily for measures that were individually administered in kindergarten and first and second grades to the subsamples of tutored poor

[2] Satz, Taylor, Friel, and Fletcher (1978) conducted an extensive longitudinal study to evaluate the precursors and correlates of reading disability and obtained evidence that visual-processing deficits as well as linguistic deficits may be found in preschoolers who later become poor readers. However, the population on which this study was based was rather heterogeneous and included a sizable number of children from impoverished environments.

readers and normal reading controls studied in-depth. Note also that the data from our kindergarten battery and additional data not reported in this article became the basis for two other studies that are discussed elsewhere, one concerned with prediction of first-grade reading achievement (Scanlon, Vellutino, Small, Chen, & Denckla, 1995) and another concerned with kindergarten instruction, relative to first-grade reading achievement (Scanlon & Vellutino, 1996).

METHOD

Participants

The initial pool of participants for this study consisted of kindergarten children from 17 schools located in six middle- to upper middle-class school districts in the Albany, New York, area. A total of 1,407 children (51% boys and 49% girls) composed the initial sample. There were two separate cohorts: Cohort 1 (n = 708) entered the study in Fall 1990 and Cohort 2 (n = 699) entered the study in Fall 1991. Before the beginning of each school year, parents of all entering kindergarten children in participating school districts were contacted and asked to involve their children in a study of reading development. We received permission from approximately 75% of the parents contacted, and the target participants of special interest in this study were initially selected from this pool.

To identify target participants, the overall reading ability of 1,284 children from the original sample of 1,407 was initially evaluated in November of first grade, at which time classroom teachers were asked to rate each child's progress in reading, writing, spelling, and math on a scale of 1 to 5, using the following criteria: A child was given a rating of 1 if he or she was having a great deal of difficulty in a given area; a rating of 2 if he or she was having some degree of difficulty; a 3 if he or she was progressing normally; a 4 if he or she was progressing somewhat more quickly than normal; and a 5 if he or she was progressing extremely well.[3] At the same time, teachers were asked to identify children who met any of the following exclusionary criteria: (a) severe vision or hearing problems, (b) frequent ear infections, (c) severe emotional problems, (d) limited intellectual ability, (e) daily medication, (f) English as a second language, and (g) diagnosed and pervasive neurological disorder. Approximately 15% of the children (125 boys and 65 girls) received a rating of 1 in reading and were considered to be potential candidates for the poor reader sample. Of these children, 4 were dropped from the study on the basis of the exclusionary criteria. The parents of all other children who received a rating of 1 in reading were contacted and asked for permission to have their children continue in the study. They were informed at that time that children who continued in the study might receive tutoring and that all participants would be periodically evaluated by means of achievement tests and various tests of cognitive ability. Ninety percent of the parents allowed their children to participate in this phase of the study.

In each classroom that ultimately contributed poor readers to the target sample, normal readers of the same sex were randomly selected from among the children in that classroom who had received teacher ratings in reading of either 3 or 4. If two poor readers of the same sex were selected from a given classroom, only one normal reader was selected from that classroom. The parents of normal readers were contacted and asked for permission to involve their children in the extended phase of the study. Additional normal readers from the same candidate pool were identified if the parents of those selected first did not agree.

All children who were potentially eligible for the target groups were given the Word Identification and Word Attack subtests of the Woodcock Reading Mastery Tests—Revised (WRMT–R; Woodcock, 1987). The Word Identification subtest evaluates facility in naming printed words individually and the Word Attack subtest evaluates the child's knowledge of letter–sound correspondences by having him or her sound out pseudowords. Because the teachers in participating school districts were widely disparate in the emphasis they placed on direct teaching of one or the other of these subskills, we endeavored to limit the selection bias that might result from different instructional histories by allowing children to qualify for the poor reader sample on the basis of performance on either the Word Identification or the Word Attack subtest. Thus, a child scoring at or below the 15th percentile on either of these measures was eligible for inclusion in the poor reader sample. To qualify for the normal reader sample, the child had to score at or above the 40th percentile on both subtests.

All children who qualified on the basis of the reading criteria were given the Wechsler Intelligence Scale for Children—Revised (WISC–R; Wechsler, 1974). To qualify, ultimately, for inclusion in the study, the child had to have an IQ at or above 90 on either the Verbal or the Performance scale of the WISC–R.[4] The number of children who ultimately qualified as poor readers was 118, which represents approximately 9% of the total population (N = 1,284) from which they were selected. There were 65 children in the normal reader control group. The poor reader group consisted of 70 boys and 48 girls. The normal reader group consisted of 36 boys and 29 girls.

Materials

Kindergarten Battery

For most measures[5] described below, reliability coefficients are presented in parentheses. Reliability coefficients for the published

[3] The 1,284 children remaining from the original sample of 1,407 reflects attrition and the failure of two teachers to return their ratings. It is worth noting that the teachers' ratings of overall reading ability were found to be quite highly correlated with measures of facility in word identification, word attack, and text reading administered toward the end of first grade (rs = .73 with word identification, .59 with word attack, and .68 with text reading).

[4] We have consistently used an IQ greater than or equal to 90 on either the Verbal or the Performance scale as the criterion for determining a child's intelligence to allow a fair evaluation of whether or not significant discrepancies between one's Verbal and Performance IQs would systematically covary with strengths and weaknesses in reading subskills, as suggested by some researchers (e.g., Kinsbourne & Warrington, 1963). Moreover, using this selection criterion in cross-sectional studies with children at or above second grade, we have found that poor readers tend to fall statistically below normal readers on the Verbal IQ, but not on the Performance IQ. Given the possibility that this pattern of results could be a consequence of cumulative deficits associated with reading difficulties (Stanovich, 1986), we were also interested in evaluating whether it would occur in beginning readers. Thus, for both reasons, it seemed prudent to use the intelligence test criterion we have used in previous research in selecting participants for this study. Finally, an additional advantage of this criterion is that it allows meaningful comparisons between given reader groups on selected subtests of the Verbal and Performance scales.

[5] In all instances, tests that do not have references indicating test publishers and publication dates, or references to researchers who constructed the tests, were experimental tests constructed at the Child Research and Study Center, University at Albany, State

tests administered are those available in the test manuals for children in the age ranges of those in this study and are presented as test–retest, internal consistency coefficients, or both. If only one coefficient is presented, it can be assumed that internal consistency was the method of estimating reliability with both published and experimental tests. Reliability coefficients for the kindergarten and first-grade experimental tasks were based on a randomly selected subsample of children tested in both kindergarten and first grade (n = 198). In cases in which no coefficients are presented for experimental tasks, it should be assumed that it was not feasible to calculate these coefficients. For example, internal consistency estimates would not be feasible in the case of speed-of-response measures or measures with a small number of items. In the case of published tests, the absence of reliability coefficients indicates that such coefficients were not provided in the test manual.

Language and Language-Based Measures

Phoneme Segmentation. This test consisted of three components administered conditionally and hierarchically: initial phoneme deletion, final phoneme deletion, and phoneme articulation. On the initial phoneme deletion component, the child was asked to say the stimulus word, delete the first sound in the word, and then say the word that remained. On the final deletion component, the child was asked to again say the word and then delete the last sound. On the phoneme articulation component, the child was asked to articulate the sounds that were different in two minimally contrasted words (e.g., *pin* and *pen*). However, because most beginning kindergartners have little experience with phoneme segmentation tasks such as these, we used an administration procedure that involved considerable modeling and corrective feedback. This was done to distinguish between children who might, with some assistance, "catch on" to the nature of speech segmentation and those who might continue to have difficulty even after modeling. These tasks are not equally difficult, and we therefore made administration of a more difficult task contingent on the child's performance on a less difficult task. To increase discriminability, we developed a scoring system that allowed for partial credit (maximum score = 24, r = .93). More detailed accounts of the administration and scoring procedures are presented in the Appendix.

Rapid Automatized Naming. This test is essentially the same as the rapid naming tests initially used by Denckla and Rudel (1976a, 1976b). It involves the presentation of a 5 × 10-in. (12.7 × 25.4 cm) array of line drawings of common objects (e.g., *book, star, hand, dog, chair*) on an $8\frac{1}{2}$- × 14-in. (21.6 × 35.6 cm) card. The child is asked to name each object in turn, and his or her score on this task is the cumulative time taken to completion as recorded with a digital stopwatch. The error score is the total number of objects misnamed (maximum error score = 50).

Rapid Articulation. We adapted this test from Stanovich, Nathan, and Zolman (1988). It entails rapid alternating repetition of word pairs in five pairs of words (e.g., *table–candy, table–candy*, etc.), and the examiner used a stopwatch to determine how long it took the child to complete seven repetitions of each pair. The performance index for this test is the mean of the times (in seconds) for each of the five word pairs.

Syntactic Processing. We used the Linguistic Concepts subtest of the Clinical Evaluation of Language Fundamentals—Revised (CELF–R; Semel, Wiig, & Secord, 1987) to measure comprehension of grammatical structures. On this subtest, the child was presented with a visual display consisting of a number of different colored lines and was instructed to point to specific lines (e.g., "Before you point to the red line, point to all of the blue lines"). The grammatical complexity of the instructions gradually increases in difficulty. Raw scores are used as the performance index on this measure (maximum score = 20; r = .85, internal consistency; r = .49, test–retest).

Semantic processing. We used the Peabody Picture Vocabulary Test—Revised (PPVT–R, Dunn & Dunn, 1981) as a measure of receptive vocabulary development. On this test, the child was presented with four pictures and was asked to point to the picture that corresponded with a word spoken by the examiner. One point was awarded for each correct response. The sum of these scores was used as the performance index (maximum score = 175, rs = .78 to .84).

Memory Measures

Sentence Memory. This test evaluated verbatim recall of orally presented sentences and consisted of seven sentences ranging in length from 4 to 12 words. One point was awarded for each sentence recalled correctly. No partial credit was given (maximum score = 7).

Word Memory. For this test, the words that comprised each of the seven sentences from the Sentence Memory Test were randomly reordered. The examiner read each of the resulting word strings to the child, and the child was asked to recall each of the words in the string in the order in which they were presented. The same random order was used for each child. One point was awarded for each random string recalled correctly (maximum score = 7).

Visual Memory. On this test, the child was presented with $8\frac{1}{2}$- × 11-in. (21.6 × 27.9 cm) matrices consisting of 9 or 12 cells. Some of the cells in each matrix contained a large black dot. These dots, collectively, formed a visual pattern. Each matrix was presented for 2 s, and the child's task was to reproduce the dot pattern from memory on a blank matrix that was drawn on a transparency that overlaid a magnetic drawing board. A round magnet was used to make dots that were the same size as those on the test stimuli. There were a total of 14 items. For four of the items, the patterns created by the dots were readily labelable (i.e., a square or a *T*); for the remaining items the patterns were not readily labelable. This feature allowed us to distinguish between visual versus verbal coding factors as possible sources of reader group differences in reproducing dot patterns. One point was awarded for each pattern that was correctly reproduced. Separate tallies were made for the labelable and nonlabelable items (maximum score = 4 labelable, 10 nonlabelable; r = .72 for both subtests combined).

Paired-associate learning. The first 57 items of the Visual-Auditory Learning subtest of the WRMT–R were administered to evaluate visual–verbal paired-associate learning of the type involved in beginning reading. This subtest presented the child with ideographs representing common words, and his or her task was to learn the names of these ideographs in order to "read" segments of text composed of them. The child was initially presented with 4 ideographs, was told the name of each, and was required to repeat each name directly after hearing it. The child was thereafter presented with several phrases constructed from the ideographs and then asked to read each phrase. Corrective feedback was given for each ideograph if the child responded incorrectly or failed to respond in 5 s. Four new ideographs were then introduced, and the child was asked to read new phrases consisting of the new as well

University of New York. Unless otherwise indicated, scoring for published tests conformed to the procedures outlined in the test manuals. Reliability coefficients for tests constructed at the center were computed by means of Cronbach's alpha.

as the old ideographs. The test continued, using the same format of introducing 4 new ideographs and then asking the child to read phrases consisting of both new and old ideographs. A total of 16 ideographs were administered in the portion of the (standardized) test administered. One point was awarded for each correct response (maximum score = 57, r = .95 for the full 134-item test).

Measures of Cognitive Processing and General World Knowledge

Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI–R; Wechsler, 1989). We used two subtests of the WPPSI–R to estimate the child's general level of intellectual functioning: the Information subtest from the Verbal scale and the Block Design subtest from the Performance scale. The Information subtest required the child to answer orally presented questions evaluating general world knowledge. The Block Design subtest required the child to use a set of colored blocks to reproduce designs that were either modeled by the examiner or presented pictorially. For both WPPSI-R subtests, raw scores are used as the performance index (maximum scores are 27 and 42 for the Information and Block Design subtests, respectively; rs = .74 to .84 for the Information subtest; rs = .79 to .86 for the Block Design subtest).

Concrete Operations. We used this task to measure three aspects of decentration ability as conceptualized by Piaget (1952): conservation, seriation, and class inclusion. Each component was measured with two different tasks that are similar to those that have appeared in the literature. Because of space constraints, these tasks are described in detail in the Appendix (maximum score = 12 summing across tasks).

Attentional and Organizational Processes

Modified Matching Familiar Figures. We used a modified version of the Matching Familiar Figures test initially used by Kagan (1965) to evaluate attentional and organizational processes. This measure evaluated the speed and accuracy with which the child could match a line drawing to one of four alternatives (Kagan used six alternatives). The test required some degree of visual analysis and attendance to subtle differences in the figures (maximum error score = 28; see the Appendix for more detail on test administration and scoring).

Target Search Test. This was a paper-and-pencil vigilance task that was initially used by Rudel, Denckla, and Broman (1978) as one of several measures evaluating what have been called "executive functions." The test presented an array of symbols, and the child was asked to locate each and every symbol that matched a target symbol, proceeding from left to right (maximum score = 18; see the Appendix for more detail on test administration and scoring for this task and for an extended version administered to target children in second grade).

Precursor and Rudimentary Reading Skills

Letter Identification. We used the Letter Identification subtest of the WRMT–R to evaluate rudimentary reading skills. It included both upper- and lowercase manuscript letters, as well as a sampling of upper- and lowercase cursive letters. The performance index used for this subtest is the raw score for the total number of letters named correctly (maximum score = 51; r = .94).

Word Identification. To identify those children who had achieved some degree of facility in reading, we administered the Word Identification subtest of the WRMT–R to evaluate decontextualized word identification. This test required oral pronunciation of "sight words" presented one at a time. The performance index used for the test is the raw score for the total number of words correctly identified (maximum score = 106; r = .98).

Word Attack. The Word Attack subtest from the WRMT–R was used to evaluate phonetic decoding ability. This subtest required that the child decode nonsense words and, thus, directly evaluated his or her knowledge of letter–sound correspondence. The performance index used for this test is the raw score for total number of nonsense syllables correctly pronounced (maximum score = 45; r = .94).

Print Awareness. This test was developed by Huba and Kontos (1985). It was designed to assess the child's understanding of the communication value of print. The measure is composed of two subtests: Picture-Choice and Picture-Word. The Picture-Choice test requires that the child select the best way to acquire a piece of information when provided with three alternatives. For example, the child might be asked to identify the best way to find out what's in a can given the following choices: "open the can, look at the label, chew a piece of gum." On the Picture-Word test, the child is asked to choose either a picture or a printed message to show the best way to convey an idea to another person. For example, the child might be asked to decide whether a picture of a seated dog or the printed sentence "My dog ran away" would be the best way to tell someone that his or her dog had run away. There are five items on each of the subtests. One point was awarded for each correct response (maximum score = 10).

Print Conventions. This test assessed the child's understanding of common print conventions in written English such as the left-to-right and top-to-bottom directionality of print, the concepts of letter and word and the meaning of punctuation marks (maximum score = 12).

Math Measures

Arithmetic subtest of the WPPSI–R. This subtest assessed comprehension of terms referring to quantitative attributes (i.e., tallest, biggest), as well as comprehension of rudimentary number concepts such as counting and one-to-one correspondence. Items at the upper end of the scale required that the child solve verbally presented arithmetic problems. The raw score is used as the performance index (maximum score = 23; r = .80).

Experimental math measure. Included in this measure were items that assessed the child's ability to (a) count by rote to 40, (b) skip count (by 2s) to 40, (c) read one-, two-, and three-digit numbers (12 items), and (d) solve written number sentences (e.g., 2 + 1 = __) that were composed of one- and two-digit addition problems (four items). Each of these skills was scored separately. The counting tasks were awarded a maximum of 10 points each. For the number reading and number sentence components, one point was awarded for each correct response.

First-Grade Selection Measures

Intellectual Ability

We used the full WISC–R to select participants for the poor and normal reader samples in first grade (for Verbal Intelligence Quotient [VIQ], Performance Intelligence Quotient [PIQ], and Full Scale Intelligence Quotient [FSIQ], rs = .91 to .95 for internal consistency; r = .90 to .94 for test–retest).

Reading Achievement

Word Identification subtest of the WRMT–R. See the *Kindergarten Battery* section.

Word Attack subtest of the WRMT–R. See the *Kindergarten Battery* section.

Basic Skills Cluster (BSC) of the WRMT–R. Scores on the Word Identification and Word Attack subtests were transformed into Rasch ability ("W") scores to obtain a composite that we used for computing percentile ranks, grade equivalents, and standard scores depicting general reading ability.

Phoneme segmentation. The Phoneme Segmentation Test administered in kindergarten was again administered on initial sample selection. Performance on this measure was not, however, used as a selection criterion.

Reading Achievement Measures Used to Evaluate Intervention Effects

Word Identification subtest of the WRMT–R. See the *Kindergarten Battery* section.

Word Attack subtest of the WRMT–R. See the *Kindergarten Battery* section.

Oral reading of connected text. This test evaluated accuracy in reading connected text orally. There were four paragraphs ranging in difficulty from the preprimer to the second-reader level. All children began reading the first paragraph and continued to read progressively more difficult paragraphs until they made more than 10 errors on a given paragraph or until the last paragraph had been administered. A maximum score of 11 points was allotted to each paragraph. The child's score for each paragraph was equal to 11 minus the number of oral reading errors (to a maximum of 11) made on that paragraph. Unadministered paragraphs were scored 0 (maximum score = 44). In a larger sample of end-of-first grade children, this measure correlated .87 with the Word Identification subtest of the WRMT–R.

Silent reading comprehension. We used the Reading Comprehension component of the Spache Diagnostic Reading Scales (DRS; Spache, 1981) to evaluate silent reading comprehension. It assessed the child's ability to read and comprehend narrative text (stories) presented in print. The score used is the grade level of the most difficult passage the child comprehended acceptably by DRS standards (maximum grade equivalent = 7.5; r = .98, internal consistency; r = .94, test–retest).

Math Achievement Measures

Calculations subtest of the Woodcock-Johnson Tests of Achievement—Revised (Woodcock & Johnson, 1989). We administered this test to assess the child's ability to perform written math calculations such as addition and subtraction (maximum score = 58; r = .93).

Applied Problems subtest of the Woodcock-Johnson Tests of Achievement—Revised. We administered this measure to assess the ability to solve math "story problems" presented auditorily (maximum score = 60; r = .84).

Cognitive Test Battery Given to Target Children in First Grade

Phonological Processing

Phoneme segmentation. This test was similar to the test administered in kindergarten and in the middle of first grade. However, twice as many items were administered (10 per subtest as opposed to 5) and the amount of modeling and feedback provided on the earlier measure was somewhat reduced on this measure. Further, no credit was given for accurate performance on the sample items (maximum score = 30, r = .95).

Phonological Memory (see also Verbal Memory). This test evaluated free recall of six auditorily presented nonsense syllables over eight trials. On each trial, items were presented randomly, and during a 6-s delay between item presentation and recall, the child was engaged in a counting task to prevent rehearsal. One point was awarded for each syllable recalled on each trial. The final score was the total number of nonsense syllables recalled correctly summed across trials (maximum score = 48).

Syntactic Processing

Token test (DiSimoni, 1978). We administered Parts IV and V of this test to evaluate the child's ability to comprehend and execute spoken directives. Each part presented a display of moveable blocks that varied in color, shape, and/or size. For each item, the examiner gave the child a spoken command, which the child was to execute by moving the blocks (e.g., "Put the red square on the blue circle"). One point was awarded for each correct response (maximum scores for Parts IV and V were 10 and 21, respectively).

Grammatic Understanding subtest of the Test of Language Development—Primary:2 (TOLD–P:2; Newcomer & Hammill, 1991). This test provided another index of sentence comprehension. For each item, the examiner presented a sentence orally, and the child chose, from among three drawings, the one that best depicted the meaning of the sentence. One point was awarded for each correct response (maximum score = 25; r = .89, internal consistency; r = .76, test–retest).

Grammaticality Judgments. This test required that the child decide whether sentences spoken by the examiner were grammatically well formed. In introducing the test, the child was given several examples of grammatically ill-formed sentences to help him or her understand the nature of the task (e.g., "The boy noticed which his dog was barking again"). One point was awarded for each correct response (maximum score = 20; r = .60).

Oral cloze. On this test, the child was given auditory presentations of short paragraphs consisting of several sentences from which given words had been deleted (Tunmer & Hoover, 1993). After hearing a practice item, the child then heard 11 test "stories," and when missing words in a given story were encountered, he or she was given 10 s to provide a plausible response. Corrective feedback was given if the child did not respond or if his or her response was incorrect. The score for this test was the total number of correct responses across sentences and paragraphs (maximum score = 25; r = .57).

Semantic Processing

Vocabulary subtest of the WISC–R. The child was presented with progressively more difficult words to define orally. A maximum of two points was awarded for each response. Scoring procedures conformed with those outlined in the test manual (Wechsler, 1974; maximum score = 64; r = .74, internal consistency; r = .68, test–retest).

Similarities subtest of the WISC–R. The child was asked to identify commonalities between pairs of objects or concepts (e.g., "In what way are a wheel and ball alike? How are they the same?"). A maximum of two points was possible for all but the early items. To be awarded maximum credit, the child had to identify a superordinate category under which both items could be classified. Scoring procedures conformed with those outlined in the test manual (maximum score = 30; r = .87, internal consistency; r = .74, test–retest).

Naming and Fluency

Rapid automatized naming. This test was described in detail in the kindergarten battery. At the first-grade level, the children were given four separate sets of items for rapid naming: objects, colors, letters, and numerals. Administration and scoring procedures were the same as those used for the kindergarten measure (maximum error score for each set = 50).

Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1983). This is a confrontational naming task that required that the child label line drawings of objects. One point was awarded for each correctly labeled object (maximum score = 60; r = .88).

Verbal fluency. We used two tasks to evaluate the speed with which the child was able to access words from permanent memory. The first, Semantic Category Fluency, was taken from the CELF–R and involved presenting the child with two categories—foods and animals—one at a time and asking him or her to name as many members of each category as he or she could think of in 1 min. The second, Phonological Category Fluency, was taken from the Controlled Word Association Test (Benton & Hamsher, 1989). It asks the child to think of as many words as possible that begin with each of three letters (*C, F,* and *L*), allotting 1 min for each letter. The score for each fluency measure was the total number of words retrieved from memory summed across the categories included on the measure. For the Phonological Fluency measure, credit was given for words having beginning consonants that sounded the same as the target consonants (e.g., *kitten* and *sit* were both counted as acceptable responses for words beginning with *C*).[6]

General Language Processing

We used the Listening Comprehension subtest of the DRS to evaluate the ability to listen to and comprehend narrative text (stories) presented auditorily. The score used was the grade level of the most difficult passage the child was able to comprehend acceptably by DRS standards (maximum grade equivalent score = 7.5; r = .94).

Verbal Memory

Digit Span subtest of the WISC–R. We used this test to evaluate short-term memory. It entails verbatim memory for randomly ordered digits, using both forward and reversed-order formats. One point was awarded for each digit string recalled correctly. Raw scores are used as the performance index (maximum score = 28; r = .76, internal consistency, r = .80, test–retest).

Sentence Imitation subtest of the TOLD–P:2. This test evaluated verbatim memory for sentences. One point was awarded for each sentence recalled correctly. Raw scores are used as the performance index (maximum score = 30; r = .92).

Recall of concrete and abstract words. This test evaluated short- and long-term verbal memory, using immediate and delayed recall formats, respectively. The immediate recall format used a selective reminding procedure (Buschke, 1973) wherein, on each trial after the first trial, only the words that were not recalled on the immediately preceding trial were presented. The stimulus list consisted of 6 concrete and 6 abstract words, and all 12 words were presented on the first trial. The immediate recall condition entailed six trials. On each trial, both types of words were randomly mixed, and recall commenced immediately after presentation of the last item. The delayed recall condition entailed only one trial and occurred approximately 20 to 25 min after the immediate recall condition. The immediate and delayed conditions were scored separately. One point was awarded for each word recalled. Separate tallies were made for concrete and abstract words (maximum scores = 36 concrete and 36 abstract on the immediate recall portion).

Phonological memory. See description in the *Phonological Processing* section.

Syntactic word order. This is an experimental test developed by Tunmer to evaluate "syntactic awareness" (Tunmer & Hoover, 1993). However, because it depends very heavily on one's ability to hold information in working memory, we used it to evaluate working memory ability. The child hears words in a sentence presented in scrambled order and must mentally rearrange the words to form a grammatically correct sentence. One point was awarded for each correct sentence (maximum score = 25; r = .74).

Visual Processing Skills

Visuomotor and visuospatial abilities. The Performance IQ was used as a composite measure of visuomotor and visuospatial ability. However, for purposes of comparison with the visual tasks used in kindergarten, we present results for the Block Design subtest separately (maximum score for Block Design = 62; r = .80, internal consistency; r = .78, test–retest).

Visual memory. On this test, the child was asked to reproduce dot patterns from memory. This measure was essentially the same as the measure administered in kindergarten except that the child was given a larger number of items and some of the dot patterns were presented using a 4 × 4 grid (maximum score = 22, 7 labelable and 15 nonlabelable; r = .72).

Procedure

The two cohorts of kindergartners were administered the test battery in 2 consecutive years. The times at which given children were tested encompassed a period extending from the summer before initiation of the school year through January of that year. The entire kindergarten test battery took approximately 2 hr and was administered in a one-to-one situation, typically in two 1-hr sessions. As indicated earlier, the poor and normal readers who served as target children were selected in mid-first grade, and the poor readers were randomly divided into two target groups: a tutored group and a nontutored group. Children in the tutored

[6] Strictly speaking, our measure of phonological fluency is not a pure measure of phonological processing because it depends, in part, on the child's familiarity with the letters *C, F,* and *L*, if not the spellings of words beginning with these sounds. However, because the child is asked to generate words in spoken language that begin with these letters, we assume that it also depends on ready access to letter names or letter sounds, both of which depend on fluency in phonological recoding, or so it would seem. Thus, although we do not deny that performance on this task is mediated, to some extent, by familiarity with a word's graphic and orthographic features, we believe that it depends more heavily on one's facility in translating these attributes into their phonological counterparts.

group (n = 76) were provided with daily one-to-one tutoring (30 min per session) for a minimum of approximately 15 weeks (typically 70 to 80 sessions). Tutoring was tailored to the child's individual needs and typically included approximately 15 min per session devoted to reading connected text. Along with facilitating reading for meaning and fun, a major objective of time spent in text reading was to foster deliberate use of a variety of strategies for word identification: sentence or thematic contexts for prediction and monitoring, external aids (e.g., picture clues), phonetic (letter–sound) decoding, and so forth. In addition, in each session, portions of time were devoted to helping the child develop a sight vocabulary, helping him or her to acquire phoneme awareness, attuning him or her to the alphabetic principle, and facilitating phonetic decoding and writing skills. The amount of time devoted to one or another of these activities was determined by the child's individual needs.

Fourteen tutors provided the remediation. All were certified in reading, elementary education, or both. All but one had at least 2 years of teaching experience before becoming involved in the project. Tutor training consisted of a 30-hr seminar supplemented by reading of theoretically and practically relevant materials. To ensure fidelity of treatment, we recorded all tutoring sessions on audiotape. For each child, one out of every 10 tutoring sessions was randomly selected to be reviewed by one of the developers of the intervention program (Frank R. Vellutino, Donna M. Scanlon, and Edward R. Sipay). These reviews served two purposes: (a) to ensure that the needs of the individual child were being met appropriately and (b) to ensure that the tutor's instructional approach was consistent with the training she had received. After reviewing the tapes, the supervisor met with individual tutors to discuss instructional issues pertinent to individual children as well as to the tutor's overall approach to instruction. In addition to these individual meetings, there were biweekly meetings of the entire group of tutors and supervisors, during which general issues of relevance to the group were discussed.

As we indicated earlier, children who did not receive tutoring (n = 42) served as a contrast group that allowed us to evaluate the relative effectiveness of our tutoring program. These children received the remediation available through their home school. However, school-based remediation varied markedly from school to school and included everything from individually tailored one-to-one remediation (that was quite similar to the remediation provided by research staff) to "small" group instruction (sometimes 9 or 10 children in a group) using a highly structured basal approach. Thus, for present purposes, the contrast group of primary interest consisted of those children who received only small-group instruction, although we present descriptive statistics for both groups. The normal reading controls did not receive any formal reading instruction except for instruction that was provided by the classroom teacher.

Between February and May of their first-grade year (during the first semester of tutoring), all target children were administered the battery of cognitive tests described in the Materials section. During June of their first-grade year (at the end of 15 weeks of tutoring), these children were given a battery of outcome measures of reading achievement that included the Word Identification and Word Attack subtests of the WRMT–R, as well as an experimental measure of oral reading accuracy constructed specifically for the purpose of assessing first-grade oral reading ability. To make a determination as to whether to continue or discontinue tutoring for a given child, as well as (initially) to assess the stability of reading status following remediation, we readministered the Word Identification and Word Attack subtests of the WRMT–R during the fall of the child's second-grade year. Those children who did not score at or above the 40th percentile on the Basic Skills Cluster (BSC) of the WRMT–R continued in remediation and were given between 8 and 10 weeks of additional tutoring. Finally, in winter and spring of their second-grade year, all target children who were not lost through attrition were again given the Word Identification and Word Attack subtests of the WRMT–R, and at spring testing, the Reading Comprehension component of the DRS was given as well.

RESULTS AND DISCUSSION

Results reported in the ensuing sections focus on the target poor readers who received daily (one-to-one) tutoring and on the normal readers. The poor readers are divided into separate groups on the basis of how well a child responded to remediation. For present purposes, response to remediation measured by the "growth rate" of the child's reading ability from initial testing in kindergarten to fall of second grade—that is, both before and shortly after remediation. Following work done by Bock (1983), Rogosa and Willett (1985), Bryk and Raudenbush (1987), and Foorman et al. (1991), the growth rates of individual children in both the tutored groups and the normal reader groups were obtained by specifying true reading status as a function of time (see also Willett, 1988). For all groups, we used Rasch-based ability (W) scores for the BSC of the WRMT–R as the estimate of reading ability at given points in time. Four measures were obtained: one in kindergarten, one during winter of first grade, one during late spring of first grade (directly after remediation), and one during fall of second grade. As indicated earlier, the BSC is a composite derived from averaging the W scores for the Word Identification and Word Attack subtests of the WRMT–R, and because the child must acquire facility in both of these subskills to learn to read, it seemed the most comprehensive and most defensible measure for purposes of evaluating growth in reading ability.

We conducted a linear regression analysis for each child separately, using time in months between fall of kindergarten and fall of second grade as the independent variable and the BSC W score as the dependent variable. The estimated slope for each child was the measure of that child's growth rate in reading ability during this period. (Slopes were derived from BSC W scores obtained only through fall of second grade to equate the tutored groups for the amount of remediation received by the children in each group, regardless of their response to remediation.) The slopes for the children were rank ordered, and the entire group was partitioned into four separate groups according to their relative status on the slopes continuum. These groups were (respectively) designated as follows: "very limited growth" (VLG), "limited growth" (LG), "good growth" (GG), and "very good growth" (VGG). (It should be noted that although 76 children were initially given daily tutoring, we subsequently lost 2 of the children through attrition. Accordingly, analyses involving all of the tutored children are based on a total of 74 participants.) To evaluate the possible effects of IQ differences on the various dependent measures, the normal reader group was partitioned into two groups by dividing it at the (normal reader) Full Scale IQ mean, yielding an average IQ normal group (AvIQNorm), and an above-average IQ normal group (AbAvIQNorm).

Given that some poor readers in the group remediated by their home schools received small group instruction and others received some degree of one-to-one tutoring (although not daily tutoring),[7] the number of participants receiving one or the other of these forms of remediation, who fell into the various achievement groups, became exceedingly small. Because of this, and because the amount of individualized instruction a child received (by virtue of pupil–teacher ratios and amount of time being tutored) was confounded with variability in approach to instruction across schools, results from these subgroups would be difficult to interpret. Thus, we present limited data on outcome measures and no data on cognitive measures for the children remediated at their home schools. However, for purposes of evaluating the relative effectiveness of our tutoring program, we separate the school-remediated group into subgroup consisting of children who received small-group instruction and children who received some amount of one-to-one tutoring at their schools and present descriptive statistics for these two groups, as well as for the group given daily tutoring by project staff. With the exception of results obtained on both the selection measures and the reading outcome measures, the data for all but the AvIQNorm readers are presented as "effect sizes," using the means and standard deviations of this group to compute standard scores. This procedure was adopted to facilitate comparisons across the various measures of interest.

Finally, to control the familywise error rate, we conducted multivariate analyses of variance (MANOVAs) followed by analysis of variance (ANOVA) and the Newman-Keuls procedure for combinations of measures that we judged, face validly, to be dependent on many of the same cognitive abilities. In most of the analyses initially conducted, only the data for three groups—VLG, VGG, and AvIQNorm—were included in the analyses (see justification below). However, additional analyses were conducted with the VLG and LG groups combined, the GG and VGG groups combined, and the AvIQNorm group. In all instances, the measures that were analyzed together by means of MANOVA are tabled adjacent to one another. All pairwise differences that are reported as significant in Newman-Keuls testing achieve at least the $p < .05$ level of significance.

Selection Measures

Table 1 presents results on the selection measures administered to target children in the winter of their first-grade year. As is evident, the children initially identified as poor readers fell well below both normal reader groups on the word identification and pseudoword decoding tests used for sample selection. In contrast, the AvIQNorm and AbAvIQNorm do not differ appreciably from one another on either of the tests. Note also that the poor readers in the respective tutored groups and the AVIQNorm readers performed at comparable levels on the intelligence measures. Thus, it is clear that, on initial sample selection, the test profiles of the poor readers chosen as our target group of tutored children satisfy the criteria typically used to define specific reading disability. It is of some significance, however, that in the tutored groups, there are substantial differences on the WRMT–R Word Identification subtest, such that those who would go on to manifest the most limited growth in reading (VLG and LG groups) had the lowest scores on this test at the outset, whereas those who would go on to manifest considerable growth in reading (GG and VGG groups) had the highest scores at the outset, $F(3, 70) = 11.20$, $p < .001$. The Newman-Keuls test revealed that both the GG and the VGG groups were statistically better than the VLG and the LG groups ($p < .05$), but neither the former groups (GG vs. VGG) nor the latter groups (VLG vs. LG) were statistically different from each other ($p > .05$). However, these groups did not differ statistically ($p > .05$) on the WRMT–R Word Attack subtest, indicating that virtually all tutored children had inadequate facility in phonetic decoding at the outset.

Initial differences between respective tutored groups, on the word identification measure used for sample selection, could have been due solely to group differences in experience and instruction. However, if this were true in all instances, then the remedial intervention program should have produced as much or more initial growth in the VLG and LG groups as in the GG and VGG groups on both the word identification test and the phonetic decoding test. Yet, the opposite pattern emerged on both tests, suggesting that group differences initially observed on the word identification test were due, in part, to the likelihood that children in the GG and VGG groups were cognitively better equipped than were children in the VLG and LG groups to learn to identify printed words. We present additional evidence in support of this suggestion later.

Outcome Measures

In this section, we focus on reading achievement in the poor readers who received daily tutoring by project staff compared with reading achievement in the normal readers. However, as a first step in documenting the relative effectiveness of the intervention program in distinguishing between difficult-to-remediate and readily remediated problem readers, Table 2 presents respective totals (and corresponding percentages) for the number of children, given either school-based remediation or tutoring by project staff, whose percentile ranks on the BSC were within given ranges at the end of first grade, after one semester of remediation. Given that the type of remediation administered in public schools typically entails group-administered instruction, the most meaningful comparison in satisfying this objective would seem to be between the children who

[7] In a few of the schools, the reading teachers consciously modeled their remedial program after the program being provided by the project's tutor. Therefore, referring to the group that did not receive remediation through the project as a "control group" would be somewhat of a misnomer.

Table 1
Scores Obtained by Tutored Poor Readers Grouped in Accord With Growth in Reading Over Time and by Normal Readers on the Reading and Intelligence Measures Administered for Sample Selection Before Intervention

	Normal readers		Tutored groups			
Measure	Average IQ (n = 28)	Above average IQ (n = 37)	VLG (n = 19)	LG (n = 18)	GG (n = 18)	VGG (n = 19)
VIQ						
M	106.14	121.51	100.89	101.11	104.11	105.42
SD	6.70	8.57	14.47	10.19	10.46	12.01
PIQ						
M	107.00	119.03	102.32	102.67	106.11	105.26
SD	9.03	5.97	9.84	9.59	13.35	9.43
FIQ						
M	106.89	122.86	101.37	101.94	105.56	105.58
SD	6.57	5.33	10.17	7.66	12.53	10.24
Word Identification raw score						
M	37.39	38.81	4.42	6.89	11.56	11.53
SD	12.91	10.83	3.34	4.59	4.62	5.51
Word Identification grade equivalent						
M	2.20	2.22	0.96	1.07	1.26	1.25
Word Attack raw score						
M	12.79	13.73	0.74	1.06	0.78	1.32
SD	8.36	7.46	2.28	1.86	0.94	1.67
Word Attack grade equivalent						
M	2.01	1.97	0.60	0.67	0.65	0.73

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; VIQ = verbal IQ; PIQ = Performance IQ; FIQ = full scale.

received daily tutoring and those who received small-group instruction. Accordingly, the school-remediated group is divided into two subgroups, one consisting of children who received small-group instruction and a second consisting of children who received some amount of tutoring (typically 3 days a week), but not daily tutoring. Because of the small number of participants in the group receiving school-based tutoring, we conducted no significance tests that included this group. However, we (cautiously) present these results for purposes of comparison.

It can be seen that in all three of these groups the largest percentage of children scored above the 30th percentile after receiving one semester of remediation. Considering that scores on the BSC, which are at or above the 30th percentile, are well within the average range (given that the 30th percentile is equivalent to a standard score of 92, with M = 100, and SD = 15), it seems reasonable to suggest that most of the children initially identified as poor readers in our sample were not "disabled" learners in the stereotypic sense of the word. Conversely, smaller percentages of children in each group scored below the 30th percentile on the BSC, and it also seems reasonable to suggest that at least some of the children who scored in this range may be "truly disabled" learners. However, it would seem, for present purposes, that the distinction between disabled and nondisabled learners could be most comfortably made in the case of children who scored in the more extreme ranges of the BSC distribution, especially those who received daily tutoring, because these children received more extensive and more individualized remediation than did children in the school-remediated groups. If, in fact, daily tutoring is potentially more effective as a "first-cut diagnostic" in distinguishing between disabled and nondisabled learners than is the type of remediation typically offered by schools (i.e., small-group instruction), then it would be expected that the daily tutoring program would place more children in the upper ranges of the BSC distribution than would small-group instruction. This pattern was, in fact, evident, but to provide statistical confirmation that it was reliable, we performed a chi-square analysis that compared children who received small-group instruction with those who received daily tutoring from project staff, in terms of the number of children scoring above the 45th percentile on the BSC versus the number scoring below the 15th percentile on this measure; the difference was found to be statistically significant, χ^2 (1, N = 58) = 4.49, $p < .05$. An additional analysis comparing these two groups in terms of the number of children scoring above or below the 45th percentile was also found to be statistically significant, χ^2 (1, N = 102) = 5.34, $p < .05$. However, a third analysis using the 30th percentile as the basis of comparison did not produce statistically significant differences between the two groups ($p > .05$), suggesting that even small-group remediation, if implemented early, can place a majority of problem readers within at least the average range of reading achievement.

These results provide some confirmation that individual

Table 2
Numbers and Percentages of Children, in Respective Remediated Groups, Whose Percentile Ranks for the Basic Skills Cluster Were Within Given Reading Achievement Ranges at the End of First Grade

Type of remediation	≤15%	16–30%	31–45%	>45%	Total
Small group instruction by school					
n	7	5	9	5	26
%	26.9	19.2	34.6	19.2	
Several days of tutoring by school					
n	4	1	4	7	16
%	25.0	6.2	25.0	43.8	
Daily tutoring by tutors trained by project staff					
n	12	13	17	34	76
%	15.8	17.1	22.4	44.7	

Note. Children are grouped by percentile ranking on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised at the end of first grade.

tutoring would, in most cases, be a more effective intervention procedure than would small-group instruction. We can, therefore, have some faith in the possibility that early and labor-intensive intervention can be reasonably effective in distinguishing between children who are difficult to remediate and those who are readily remediated and, thus, between children who might be classified as disabled learners, despite even optimal intervention, and those who need not be so classified given adequate intervention. In fact, even the limited amount of tutoring provided by school personnel placed a higher percentage of children above the 45th percentile on the BSC than did the small-group instruction provided by school personnel, although the percentage of children who scored below the 15th percentile on the BSC was no less in the group-provided school-based tutoring than in the group-provided small-group instruction.

That our tutoring program was reasonably successful in distinguishing between children who are difficult to remediate and those who are more readily remediated is given additional confirmation from the results of the growth curve analyses comparing the progress of children in respective tutored groups relative to that of the normal readers, on measures of word identification and phonetic decoding. (Although the tutored groups were partitioned on the basis of slopes for BSC W scores obtained from kindergarten through fall of second grade, we present data for winter and spring of second grade to demonstrate stability in the rank ordering of these respective groups following remediation.) Figure 1 presents growth curves for raw scores on the WRMT–R Word Identification subtest. Focusing initially on the measures administered before initiation of the intervention program (kindergarten to Grade 1 winter), we see that the gradients for growth in the AvIQNorm and AbAvIQNorm groups are virtually identical and are much steeper than those of the tutored children, reflecting the finding that the children in both normal reader groups initially had much greater success in learning to identify printed words than did the children in the tutored groups. However, between winter and spring of first grade, after only one semester of remediation, there was a dramatic increase in the performance levels of each of the tutored groups, attesting to the positive effects of remediation. Note also that the gradients for this period are steeper in the GG and VGG groups than in the VLG and LG groups, reflecting differences between these respective groups in their initial response to remediation. The gain scores corresponding with these differences are presented in Table 3. (For purposes of comparison, we present the gain scores for the normal readers during the same period.) An ANOVA evaluating gain score differences in the tutored groups yielded a statistically significant effect, $F(3, 70) = 17.91$, $p < .001$, and post hoc testing by means of the Tukey-Hayter procedure (Seaman, Levin, & Serlin, 1991) revealed that gain scores in the GG and VGG groups were each statistically larger ($p < .05$) than gain scores in the VLG and LG groups. And whereas gain scores in the latter groups were not statistically different from each other ($p > .05$), gain scores in the VGG group were statistically larger ($p < .05$) than gain scores in the GG group.

From the winter of first grade to the spring of second grade, both the AvIQNorm and the AbAvIQNorm groups were on a gradually rising and identical growth trajectory that was (roughly) paralleled by the tutored groups from spring of first grade to spring of second grade (except for the period between Grade 1 spring and Grade 2 fall, during which the VLG, the LG, and GG groups dropped somewhat or leveled off, while the VGG continued to rise somewhat). An ANOVA comparing the slopes indexing growth rates for the tutored groups and the normal readers during this period yielded null results ($p > .05$), indicating that growth rates in these respective groups were comparable after intervention (see Table 3). However, each of the tutored groups maintained its relative status over the period evaluated, suggesting that initial response to remediation may be a reasonably stable indicant of whether or not a problem reader would continue to have significant difficulty in learning to read. This suggestion is given additional support by results on the test of phonetic decoding ability.

Assessment of change in phonetic decoding ability produced a pattern of results similar to the pattern noted above, but with some important differences. Figure 2 presents the growth curves based on raw scores for the Word Attack (pseudoword decoding) subtest of the WRMT–R. It can be seen that the pattern of change for respective tutored groups is similar to that for growth in word identification, insofar as the children in these groups made little or no progress in phonetic decoding ability before receiving any remediation. However, the growth rates in these groups increased sharply from the winter of first grade to the spring of first grade, after one semester of remediation. As in the case of word identification, the gradients representing growth in phonetic decoding ability from Grade 1 winter to Grade 1 spring are steeper in the GG and VGG groups than in the VLG and LG groups. The gain scores corresponding with these gradients

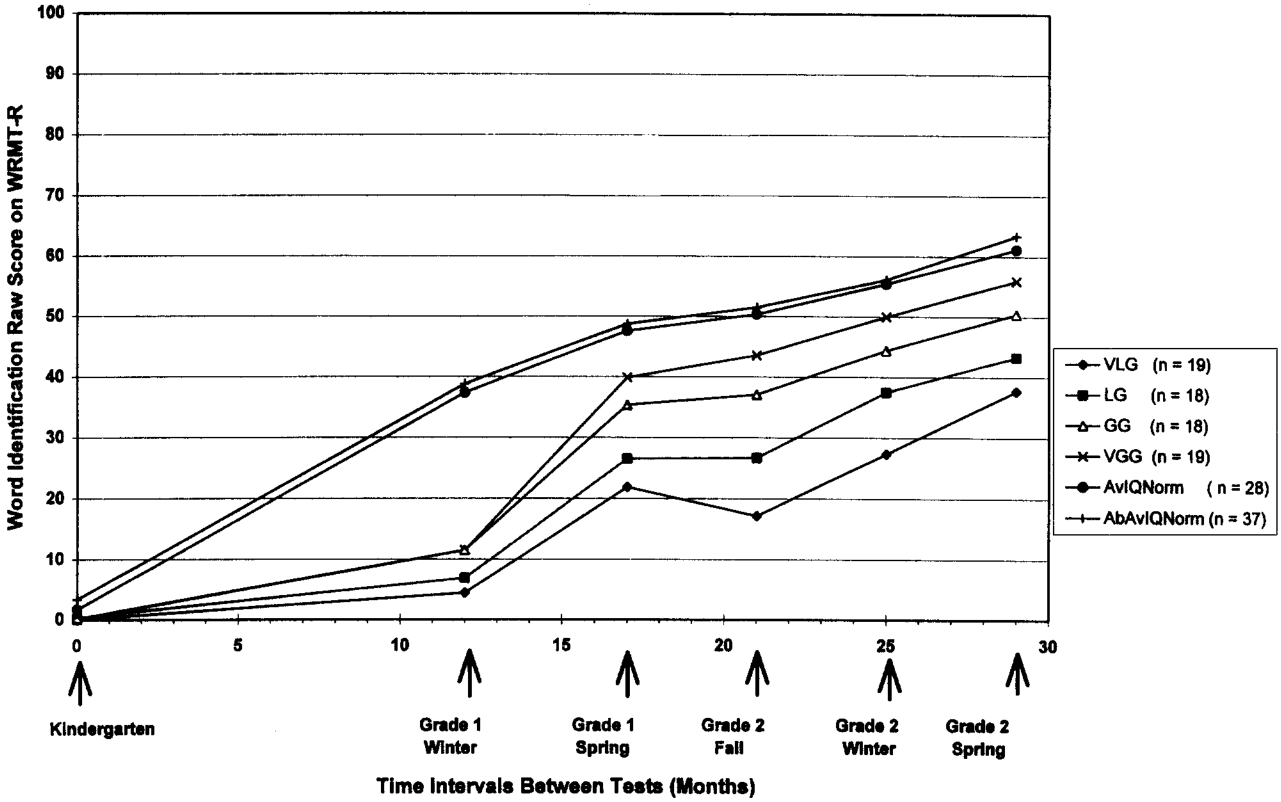

Figure 1. Growth curves for raw scores on the Woodcock Reading Mastery Tests—Revised (WRMT–R) Word Identification subtest for normal readers and tutored poor readers. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; AvIQ Norm = average IQ normal group; AbAvIQNorm = above average IQ normal group.

were found to be statistically different, $F(3, 70) = 18.42$, $p < .001$, and post hoc Tukey-Hayter testing again revealed that (a) gain scores in the VGG and the GG groups were statistically larger ($p < .05$) than gain scores in the VLG and LG groups, (b) gain scores in these latter groups were not statistically different from each other ($p > .05$), and (c) gain scores in the VGG group were statistically larger ($p < .05$) than gain scores in the GG group.

Growth rates in the AvIQNorm and AbAvIQNorm groups increased sharply from kindergarten to winter of first grade and continued to increase steadily thereafter. And, as was true of word identification, the growth rates of these two groups were virtually identical. However, as is evident in Figure 2, phonetic decoding in the VGG group increased sharply during the intervention period and their performance became more like that of the normal readers after intervention. The slopes representing growth in this ability over the period encompassing Grade 1 spring to Grade 2 spring are presented in Table 3. Slope differences among the tutored groups for this period were found to be statistically significant, $F(3, 70) = 4.04$, $p < .05$. The Tukey-Hayter test revealed that the slopes for the VGG, GG, and LG groups were not statistically different from each other ($p > .05$), but the slopes for the VGG and LG groups were statistically larger ($p < .05$) than the slope for the VLG group. Slope differences were also found to be statistically significant when the normal readers were included in the analysis, $F(5, 128) = 2.26$, $p = .05$, but the only difference that was found to be statistically significant in post hoc Tukey-Hayter testing was that between the VGG and the VLG groups.[8] Thus, it should be clear from both their ultimate level of functioning and their rate of growth following remediation that children in the VGG group were indeed closer to normal readers than to children in the other tutored groups in phonetic decoding ability and that, among the tutored groups, the greatest disparity that emerged in this ability as a consequence of remediation was that between the VGG and the VLG groups. Given the heavy reliance of phonetic

[8] The number of children included in the slopes analyses is reduced by four for Word Identification and by five for Word Attack. Two children in the AvIQNorm group and 1 child in the AbAvIQNorm group were eliminated from these analyses because they had left the study before data collection was complete, and as a result, slopes for the period between spring of first grade and spring of second grade could not be computed. Two additional children were eliminated in the AbAvIQNorm group, 1 because her kindergarten Word Identification and Word Attack scores were unusually high (which suggests she was not a "normal" reader) and another because of missing data on the Word Attack subtest.

Table 3
Gain Scores Depicting Growth in Reading Subskills From Winter of First Grade to Spring of First Grade and Slopes Depicting Growth in Reading Subskills From Spring of First Grade to Spring of Second Grade

	Normal readers		Tutored groups			
Reading subskill	Average IQ (n = 28)	Above average IQ (n = 37)	VLG (n = 19)	LG (n = 18)	GG (n = 18)	VGG (n = 19)
	Gain scores for winter to spring of first grade					
Word Identification						
M	10.21	9.81	17.47	19.67	23.83	28.37
SD	7.11	6.63	5.94	3.58	5.39	4.46
Word Attack						
M	6.29	4.75	4.58	6.28	9.72	13.63
SD	7.74	5.86	3.17	2.95	3.66	5.75
	Slopes for spring of first grade to spring of second grade					
Word Identification						
M	1.09	1.12	1.33	1.37	1.21	1.27
SD	0.48	0.38	0.60	0.32	0.36	0.42
Word Attack						
M	0.53	0.56	0.23	0.58	0.52	0.64
SD	0.52	0.40	0.36	0.37	0.46	0.39

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth.

decoding on phonological coding ability, these results lead one to expect that the VGG group should also be closer to normal readers than to the other three tutored groups and farthest from the VLG group on nonreading cognitive tasks that also depend on phonological coding ability.

For purposes of comparison, Figure 3 presents growth curves for W scores on the BSC. As we indicated earlier, BSC W scores combine results from the Word Identification and Word Attack subtests of the WRMT–R, and they can, therefore, be considered reasonable estimates of one's ability to learn to read in the general sense. However, the differences in respective growth patterns of the tutored groups depicted in Figure 3 are more conventionally depicted in terms of percentile ranks on the BSC in Table 4. Table 4 also presents (a) raw score means and standard deviations for the oral (text) reading test administered to the children in each of these groups at the end of first grade and (b) grade equivalents on the test of reading comprehension administered at the end of second grade. It can be seen that the rank orderings of (respective) group means on the BSC obtained in first and second grades are cross-validated by the rank orderings of group means on both the text reading and reading comprehension tests, both of which provide independent measures of reading ability. Thus, given that the children in the VGG group manifested the greatest amount of progress in reading subskills over the period evaluated, whereas children in the VLG group manifested the least amount of progress, we suggest that the strongest test of the hypothesis that specific reading disability, in some cases, may be caused by basic deficits in reading-related cognitive abilities would be provided by comparison of these two groups, relative to the normal readers, on the tasks designed to evaluate these abilities. Thus, analyses conducted initially involve only the VLG, VGG, and AvIQNorm groups. However, as we indicated earlier, additional analysis compares the VLG and LG groups combined with the GG and VGG groups combined, relative to the AvIQNorm group.

Precursor and Rudimentary Skills

Table 5 presents effect sizes derived from the various kindergarten measures evaluating precursor and rudimentary literacy skills, as well as math skills and conceptual development. It can be seen that the children in the various tutored groups generally fell below the normal readers on most of these measures. However, in the analysis comparing the groups of special interest (i.e., the VLG, VGG, and AvIQNorm groups) differences were found to be statistically significant only in the case of tests evaluating rudimentary reading and math skills: Rao's $F(4, 124) = 14.70$, $p < .01$, for Letter and Word Identification; Rao's $F(18, 118) = 4.87$, $p < .01$, for WPPSI–R Arithmetic, Counting by 1s and 2s, and Number Identification. Newman-Keuls tests revealed statistically significant differences between the AvIQNorm and both the VGG and the VLG groups on most of these measures and between the VGG and the VLG groups on Letter Identification, Number Identification, and Counting by 1s.

These results make it clear that some, but not all, reading-related kindergarten skills were deficient in children who were initially identified as poor readers. Given that the children in the tutored groups performed well below the normal readers in learning letter names, it is not surprising that they also performed significantly below the normal

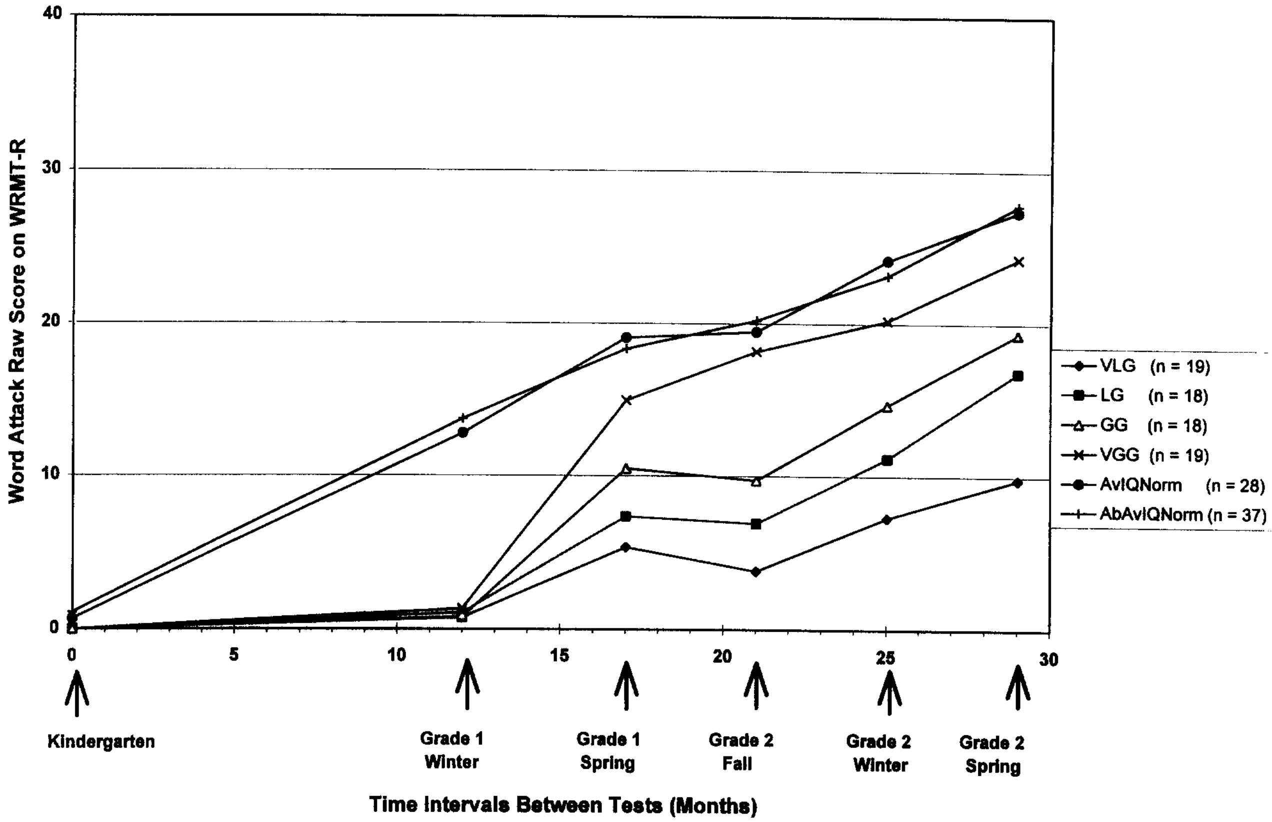

Figure 2. Growth curves for raw scores on the Woodcock Reading Mastery Test—Revised (WRMT–R) Word Attack subtest for normal readers and tutored poor readers. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; AvIQNorm = average IQ normal group; AbAvIQNorm = above average IQ normal group.

readers in acquiring an initial corpus of sight words. Deficiencies in acquiring these skills could be due either to inadequate experience or to more basic difficulties in name encoding, name retrieval, and visual–verbal association learning, although these sources of difficulty are not mutually exclusive. It is, therefore, of some significance that the best and worst achieving tutored groups were not found to be substantially different from normal readers, or from each other, on the measures evaluating basic literacy (print) concepts and conceptual development administered in kindergarten. Because the latter findings imply that most of the children in the tutored groups had adequate exposure to literacy concepts and were not generally impaired in learning, the pattern of results generated by the rudimentary reading measures could be taken as an indication that at least some of the children in each of the tutored groups had some basic difficulties learning the names of things. Reinforcing this possibility is the finding that children in the VLG group were also found to be less proficient than the normal readers and children in the VGG group on tests evaluating rudimentary counting and number identification skills. Counting entails name encoding and name retrieval, whereas number identification entails visual–verbal learning as well as name encoding and name retrieval, and these are basic cognitive abilities that underlie the acquisition of rudimentary reading abilities and, apparently, the acquisition of rudimentary math abilities as well.

Additional support for our suggestion that most of the children initially identified as poor readers were not generally impaired in learning is provided by our finding that the tutored groups were not statistically different from the AvIQNorm group (or from each other) on the intelligence measures administered in first grade. Still more support comes from the finding that the scores achieved by the tutored children on the math measures administered in first grade are all within the average range (34.69th percentile to 55.91st percentile). However, the normal readers generally performed better than the tutored children on these measures, Rao's $F(4, 124) = 3.05$, $p < .05$. Although group differences were found to be statistically significant ($p < .05$) only in contrasts between the AvIQNorm and VLG groups, it is clear that the normal readers were somewhat better in math than were the children in the tutored groups, despite the fact that math abilities in the latter groups were well within normal limits.

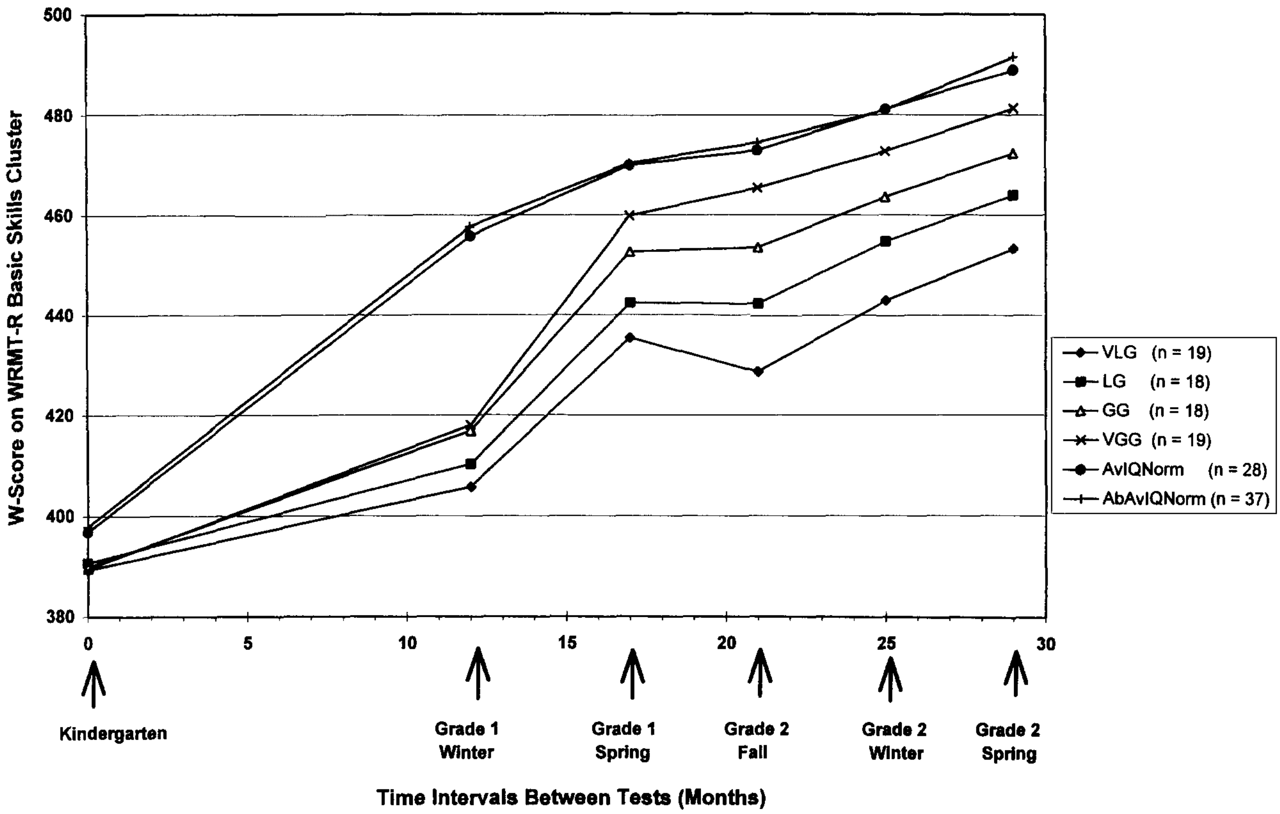

Figure 3. Growth curves for W scores on the Woodcock Reading Mastery Test—Revised (WRMT–R) Basic Skills Cluster for normal readers and tutored poor readers. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; AvIQNorm = average IQ normal group; AbAvIQNorm = above average IQ normal group.

Cognitive Abilities Underlying Reading Ability

Phoneme Segmentation

Table 6 presents the data for children in respective reader groups on the phoneme segmentation tasks administered in kindergarten and first grade. Once again, the children in the tutored groups generally fell below the normal readers on these tasks at both kindergarten and first-grade testing. However, in the analyses involving the VLG, VGG, and AvIQNorm groups, observed differences were found to be statistically significant only at first-grade testing, Rao's $F(2, 60) = 2.53, p > .05$, kindergarten; Rao's $F(4, 124) = 11.67$, $p < .01$, first grade. Post hoc testing revealed significant differences ($p < .05$) between the AvIQNorm group and both of the tutored groups, as well as between the VGG and the VLG groups in the winter of first grade. In contrast, significant differences emerged only between the AvIQNorm and the VLG groups and between the VGG and the VLG groups in spring of first grade, after one semester of remediation. These results are consistent with the possibility that reader group differences that are typically observed on phoneme segmentation tasks are due not only to experience in reading and spelling (Ehri, 1989; Perfetti, Beck, Bell, & Hughes, 1987) but also to phonological coding deficits. The only contraindication to this interpretation is the failure to observe statistically significant group differences on the test of phoneme segmentation administered in kindergarten, before the children had any extensive literacy instruction. However, in analyses that we discuss in fuller detail below, we compare the AvIQNorm group with high and low reading growth groups, respectively, consisting of the VGG and GG groups combined ($n = 37$) and the VLG and LG groups combined ($n = 37$), and in this analysis, statistically significant differences did emerge on the phoneme segmentation measure administered in kindergarten, $F(2, 96) = 3.97$, $p < .05$. Newman-Keuls testing revealed that the AvIQNorm group was statistically better ($p < .05$) than the low-reading-growth group on this measure, but they were not statistically better ($p > .05$) than the high-reading-growth group. However, the high- and low-growth groups were not statistically different ($p > .05$) from each other. Thus, results from the initial analysis of the kindergarten data may have been due to lack of statistical power, and we, therefore, conclude that observed differences between poor and normal readers on phoneme seg-

Table 4
Reading Achievement Outcome Measures Administered to Normal Readers and Children in Respective Tutored Groups During the Winter and Spring of Their First Grade Year and During the Fall, Winter, and Spring of Their Second Grade Year

	Normal readers		Tutored groups			
Grade, time of year, and measure	Average IQ (n = 28)	Above average IQ (n = 37)	VLG (n = 19)	LG (n = 18)	GG (n = 18)	VGG (n = 19)
First grade (winter)						
Basic Skills Cluster Percentile Rank						
M	72.21	72.65	8.58	13.56	18.72	22.26
SD	16.68	16.99	9.99	8.15	6.99	10.92
First grade (spring)						
Basic Skills Cluster Percentile Rank						
M	74.21	76.11	19.84	29.11	47.83	62.63
SD	22.19	19.13	12.47	8.31	9.78	11.99
Text Reading (Raw score)						
M	33.18	33.72	6.32	13.74	22.00	28.58
SD	9.16	8.49	5.02	7.83	7.33	6.18
Second grade (fall)						
Basic Skills Cluster Percentile Rank						
M	70.35	73.69	5.47	17.11	35.72	60.16
SD	24.04	23.38	3.29	5.75	8.15	13.49
Second grade (winter)						
Basic Skills Cluster Percentile Rank						
M	72.31	73.42	9.00	22.56	38.83	58.21
SD	20.84	21.03	7.92	11.06	14.65	17.64
Second grade (spring)						
Basic Skills Cluster Percentile Rank						
M	74.62	79.11	14.37	27.56	43.50	63.79
SD	20.57	15.96	16.66	12.12	16.83	15.08
Reading Comprehension Grade Equivalent						
M	3.84	4.81	2.03	2.81	3.27	3.31

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth.

mentation tasks are due, in part, to phonological coding deficits. Results on several other language-based tasks included in our test battery provide independent evidence that some of the tutored children may have been impaired by such deficits.

Naming and Verbal Fluency

Table 7 presents results on the naming and verbal fluency tasks administered to target children. The first finding of note is that the children in the tutored groups generally performed below the normal readers on the rapid automatized naming and rapid articulation tasks administered in kindergarten (positive effect sizes signify more time taken to respond and a larger number of errors, respectively). This is reflected in significant main effects for the measures (respectively) evaluating speed of object naming, $F(2, 60) = 3.29$, $p < .05$, Object Naming Time, and facility in speech articulation, $F(2, 60) = 10.19$, $p < .001$, Rapid Articulation. Post hoc testing revealed statistically significant differences on these measures ($p < .05$), but only between the AvIQNorm and VLG groups. There were no statistically significant differences, $F(2, 60) = 2.80$, $p > .05$, on the measure evaluating accuracy in object naming. On the rapid naming tasks administered in first grade, significant group differences again emerged only on the speed of response tasks, Rao's $F(8, 120) = 4.96$, $p < .01$. In this instance, however, we observed group differences on contrasts between the AvIQNorm and VLG groups and between the VGG and VLG groups, but not between the AvIQNorm and VGG groups. Significant reader group differences also emerged on the tests evaluating confrontational naming (Boston Naming Test), $F(2, 63) = 9.48$, $p < .01$, but only between the AvIQNorm and VLG groups and

Table 5
Measures of Rudimentary Reading, Math, General Knowledge, and Conceptual Skills Administered to Normal Readers and Children in Respective Tutored Groups in Kindergarten and First Grade

Measure	Normal readers: Average IQ (*n* = 28)	Normal readers: Above average IQ (*n* = 37)	Tutored groups: VLG (*n* = 19)	LG (*n* = 18)	GG (*n* = 18)	VGG (*n* = 19)
		Kindergarten				
Rudimentary reading skills						
Letter identification[a,b,c]						
M	27.57	0.11	−2.93	−2.17	−1.73	−1.52
SD	5.69					
Word identification[a,b]						
M	1.71	0.62	−0.62	−0.52	−0.58	−0.60
SD	2.68					
Concepts of print						
Print awareness						
M	6.71	0.50	−0.33	−0.23	−0.16	0.10
SD	2.35					
Print conventions						
M	11.18	0.04	−0.56	−0.61	−0.68	−0.07
SD	2.39					
Rudimentary math skills						
Counting by 1s[a,c]						
M	7.89	0.15	−1.20	−1.00	−1.00	−0.45
SD	2.28					
Counting by 2s						
M	1.11	0.05	−0.38	−0.34	−0.40	−0.38
SD	1.85					
Number identification[a,b,c]						
M	9.48	−0.06	−2.31	−1.67	−1.02	−1.24
SD	1.65					
Arithmetic (WPPSI–R)[a,b]						
M	15.61	0.48	−1.07	−0.95	−0.85	−0.70
SD	2.74					
Conceptual development						
General Knowledge (WPPSI–R Info)						
M	21.00	0.60	−0.65	−0.08	−0.08	−0.42
SD	2.02					
Concrete operations						
M	6.56	0.16	−0.66	−0.29	−0.47	−0.38
SD	3.85					
		First grade				
Math						
WJ calculation[a]						
M	8.32	0.45	−0.80	−0.29	−0.29	−0.42
SD	2.25					
WJ Applied problem[a]						
M	25.14	0.40	−0.77	−0.64	−0.36	−0.44
SD	4.86					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the AbAvIQNorm reader group. In this and all other tables, all pairwise comparisons achieve at least a p value of $\leq .05$ on the basis of post hoc testing using the Newmans-Keuls procedure. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; WPPSI–R = Wechsler Preschool and Primary Scale of Intelligence—Revised; AbAvIQNorm = above average IQ normal, AvIQNorm = average IQ normal; WJ = Woodcock-Johnson Psychoeducational Battery.
[a] Significant differences between AvIQNorm group versus the VLG group. [b] Significant differences between AvIQNorm group versus the VGG group. [c] Significant differences between VGG group versus the VLG group.

Table 6
Phoneme Awareness Measures Administered to Normal Readers and Children in Respective Tutored Groups in Kindergarten and First Grade

Measure	Normal readers: Average IQ (n = 28)	Normal readers: Above average IQ (n = 37)	Tutored groups: VLG (n = 19)	Tutored groups: LG (n = 18)	Tutored groups: GG (n = 18)	Tutored groups: VGG (n = 19)
		Kindergarten				
Phoneme segmentation						
M	5.08	0.07	−0.53	−0.49	−0.43	−0.26
SD	7.18					
		First grade				
Phoneme segmentation						
Winter[a,b,c]						
M	9.63	0.23	−1.59	−0.86	−1.28	−1.08
SD	4.88					
Spring[a,c]						
M	21.64	0.37	−1.36	−0.57	−1.24	−0.59
SD	7.23					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the above average IQ normal reader group. AvIQNorm = average IQ normal; VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth.
[a] Significant differences between the AvIQNorm group versus the VLG group. [b] Significant differences between the AvIQNorm group versus the VGG group. [c] Significant differences between the VGG group versus the VLG group.

between the VGG and VLG groups. However, group differences on the tests evaluating semantic category fluency and phonological category fluency (respectively) did not achieve statistical significance, both $ps > .05$.

These results are generally consistent with results obtained elsewhere (Blachman, 1984; Denckla & Rudel, 1976a, 1976b; Donahue, 1986; Vellutino et al., 1995; Wolf, 1984) and, taken together, provide strong support for the possibility that phonologically based fluency deficits are causally related to reading disability. If performance on rapid naming and confrontational naming tasks are determined in significant measure by one's ability to activate and retrieve phonological codes from memory with fluency and dispatch, then the finding of strong reader group differences on these tasks can be taken as evidence that some impaired readers are encumbered by basic deficits in phonological coding that directly affect name encoding and name retrieval as reading subskills. Moreover, given that we observed significant differences on many of these tasks, not only between the normal readers and the worst achieving tutored group, but also between the best and the worst achieving tutored groups and not between the normal readers and the best achieving tutored group, we can have more faith in the suggestion that the early reading problems of children in the best achieving tutored group were not caused primarily by basic deficits in phonological abilities underlying reading ability. The only contraindication to these inferences is the failure to find a similar pattern of results on the phonological fluency task. It will be recalled that the phonological fluency task required that children generate as many words as they could think of that begin with the letters *C*, *F*, and *L*, giving 1 min for each letter. This task entails familiarity with each of these letters, familiarity with words that begin with the sound of each, and name retrieval; it is not clear why it did not discriminate between respective groups. Less surprising is the absence of reader group differences on the semantic fluency task. This task required that children generate as many words as they could think of in a given semantic category (e.g., animals) and the fact that it failed to produce reader group differences could be taken as evidence that poor readers of the type studied herein are not impaired by semantic deficits that might affect name encoding and name retrieval. We earlier discussed results that are consistent with this possibility (Vellutino & Scanlon, 1987a; Vellutino et al., 1988; Vellutino et al., 1995). Additional support is provided by results on other semantic tasks administered in our test battery, which we discuss momentarily.

Finally, the finding of significant differences between the VLG and AvIQNorm groups on the rapid articulation task raises the question of whether speech–motor programming may covary in some way with naming speed. To evaluate this possibility, we computed the correlations between the rapid articulation and the object naming tasks administered in kindergarten for the tutored children and the normal readers separately and found that whereas these two variables were significantly correlated in the tutored children ($r = .31, p < .01$), they were not significantly correlated in the normal readers ($r = .18, p > .05$). Moreover, we did not find rapid articulation to be significantly correlated with growth in reading ability in either the tutored children or the

Table 7
Naming and Verbal Fluency Measures Administered to Normal Readers and Respective Tutored Groups in Kindergarten and First Grade

Measure	Normal readers: Average IQ (*n* = 28)	Normal readers: Above average IQ (*n* = 37)	Tutored groups: VLG (*n* = 19)	Tutored groups: LG (*n* = 18)	Tutored groups: GG (*n* = 18)	Tutored groups: VGG (*n* = 19)
	Kindergarten					
RAN Objects time[a]						
M	68.98	−0.15	0.84	0.69	0.67	0.53
SD	16.69					
RAN Objects errors						
M	1.04	0.37	0.75	0.91	0.79	0.49
SD	1.43					
Rapid Articulation[a] time						
M	6.95	0.19	1.90	0.74	1.26	0.70
SD	1.35					
	First grade					
RAN Objects time[a]						
M	55.38	−0.14	0.93	0.40	0.49	0.64
SD	12.04					
RAN Colors time[a]						
M	52.93	−0.18	1.53	0.15	0.53	0.87
SD	8.60					
RAN Letters time[a,b]						
M	37.36	−0.19	1.44	1.10	0.90	0.04
SD	10.80					
RAN Numbers time[a,b]						
M	40.83	−0.37	0.79	0.54	0.39	−0.19
SD	12.25					
RAN Objects errors						
M	1.32	−0.32	0.38	0.27	0.21	−0.15
SD	1.09					
RAN Colors errors						
M	1.00	−0.16	0.44	0.05	0.56	0.27
SD	1.19					
RAN Letters errors						
M	0.57	0.20	2.07	2.09	0.75	0.34
SD	0.79					
RAN Numbers errors						
M	1.18	−0.20	−0.01	−0.03	−0.11	−0.22
SD	2.00					
Boston Naming Test correct[a,b]						
M	35.39	0.41	−1.35	−0.80	−0.82	−0.54
SD	5.79					
Semantic Category Fluency total						
M	22.07	0.35	−0.29	−0.19	−0.33	−0.01
SD	7.08					
Phonological Category Fluency total						
M	14.00	0.39	−0.30	−0.28	0.10	−0.11
SD	7.28					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the above average IQ normal reader group. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; RAN = rapid automatized naming (time reported in seconds; errors reported as total across 50 items).
[a] Significant difference between average IQ normal readers versus VLG tutored readers. [b] Significant difference between VGG versus VLG tutored readers.

normal readers, whereas we found object naming as well as other rapid naming variables to be significantly correlated with growth in reading ability. Stanovich et al. (1988) obtained similar results and asserted that although articulation speed may serve as a "marker variable for phonological problems at deeper levels" (p. 72), it is doubtful that it is causally related to reading disability (see also Catts, 1986). Our data are consistent with this assertion.

Semantic, Syntactic, and General Language Processing

Table 8 presents the data for several measures that differentially depend on the semantic and syntactic components of language. We should first point out that significance tests conducted for the PPVT–R administered in kindergarten yielded null results ($p > .05$), as did those conducted for the WISC–R Vocabulary and Similarities tests administered in first grade. Because performance on each of these tests depends heavily on semantic concept development, as does performance on the semantic fluency task, we are at liberty to suggest, in accord with previous findings (Vellutino & Scanlon, 1987a; Vellutino et al., 1988, Vellutino et al., 1995), that semantic deficits are not significant causes of difficulty in learning to read, at least not in beginning readers of the type evaluated in this study.

Results on the measures that depend more heavily on

Table 8
Semantic, Syntactic, and General Language Measures Administered to Normal Readers and Respective Tutored Groups in Kindergarten and First Grade

	Normal readers		Tutored groups			
Measure	Average IQ (n = 28)	Above average IQ (n = 37)	VLG (n = 19)	LG (n = 18)	GG (n = 18)	VGG (n = 19)
		Kindergarten				
PPVT–R						
M	67.18	0.58	−0.43	−0.20	−0.53	−0.13
SD	11.35					
CELF–R Linguistic Concepts						
M	15.79	0.35	−0.80	−0.21	−0.77	−0.55
SD	2.96					
		First grade				
WISC–R Vocabulary						
M	22.64	0.58	−0.52	−0.24	−0.55	0.03
SD	6.88					
WISC–R Similarities						
M	8.86	1.12	−0.33	−0.35	−0.06	−0.38
SD	2.27					
Token Sentence Comprehension (IV)[a,b]						
M	8.00	0.36	−1.94	−0.81	−1.35	−1.39
SD	1.44					
Token Sentence Comprehension (V)[a,c]						
M	16.07	0.34	−1.13	−0.41	−0.14	−0.23
SD	2.85					
Grammaticality Judgments						
M	11.64	−0.01	−0.24	−0.20	−0.11	−0.21
SD	2.87					
TOLD–P:2 Grammatic Understanding						
M	20.96	0.44	−0.51	−0.43	−0.11	−0.38
SD	2.12					
Oral Cloze[b]						
M	17.14	0.64	−0.34	−0.81	−0.33	−0.91
SD	2.77					
Listening Comprehension[a,b]						
M	40.14	0.66	−0.60	−0.63	−0.57	−1.03
SD	11.86					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the AbAvIQNorm reader group. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; PPVT–R = Peabody Picture Vocabulary Test—Revised; CELF = Clinical Evaluation of Language Fundamentals—Revised; WISC–R = Wechsler Intelligence Scale for Children—Revised; TOLD–P:2 = Test of Language Development— Primary:2; AbAvIQNorm = above average IQ normal; AvIQNorm = average IQ normal.
[a] Significant difference between AvIQNorm versus VLG group. [b] Significant difference between the AvIQNorm group versus the VGG group. [c] Significant difference between the VGG group versus the VLG group.

syntactic abilities are mixed. No significant differences emerged ($p > .05$), either on the test evaluating the child's ability to comprehend grammatical structures administered in kindergarten (CELF–R Linguistic Concepts) or on the tests (respectively) evaluating sentence comprehension (TOLD–P:2 Grammatic Understanding) and the ability to detect grammatically ill-formed sentences (Grammaticality Judgments) administered in first grade. However, on two other measures designed to evaluate sentence comprehension administered in first grade—that is, Token Test Parts IV and V—group differences did emerge, Rao's $F(4, 124) = 7.88, p < .01$. Newman-Keuls tests revealed that the AvIQNorm group performed significantly better ($p < .05$) than both the VLG and the VGG groups on Token Part IV. They also performed better than the VLG group on Token Part V, but not better than the VGG group. Moreover, the VGG group was found to be significantly better than the VLG group on Token Part V. We also found the F ratio for group differences on the oral cloze task (inserting missing words in incomplete sentences) to be statistically significant, $F(2, 63) = 3.53, p < .05$, but follow-up tests yielded significant differences only between the AvIQNorm and the VGG group on this task.

Finally, reader group differences emerged on the test evaluating general language comprehension (DRS Listening Comprehension), $F(2, 63) = 7.57, p < .01$, but group differences were statistically significant ($p < .05$) only in contrasts between the AvIQNorm group and both the VLG and the VGG groups, respectively.

Although the finding of statistically significant differences between the tutored children and the normal readers on certain of the syntactic and general language measures could be interpreted to mean that children in the tutored groups were impaired by syntactic or general language deficits that contributed, directly or indirectly, to their difficulties in learning to read, the lack of consistency across measures compromises this interpretation. Moreover, the fact remains that null findings emerged in reader group contrasts on most of these measures, including one administered before the tutoring program was initiated (CELF–R Linguistics Concepts). One possible reason for the inconsistency observed on the syntactic measures is that some of them may have made greater demands on working memory than others. This may be especially true of Parts IV and V of the Token Test, both of which make heavy demands on working memory, by virtue of the fact that performance depends greatly on storage and accurate recall of increasingly long and complicated series of orally presented directives depicting operations that must be performed in tandem (e.g., "Touch the small white square with the large red square"). This is especially true of Part IV. The listening comprehension test and the oral cloze test also make heavy demands on working memory, because each requires that one hold information in short-term storage long enough to compute and integrate the semantic representations necessary for comprehending sentences. It is, therefore, possible that the finding of differences between AvIQNorm and the tutored groups on each of these measures could be due, in part, to group differences in working memory capacity associated with phonological coding deficits, as suggested by some researchers (e.g., Shankweiler et al., 1992; see earlier discussion).

Verbal Memory and Verbal Learning

That working memory limitations associated with phonological coding deficits may be characteristic of poor readers is given more direct support by results obtained on the verbal memory and verbal learning measures administered to target children in kindergarten and first grade. Table 9 presents the data for these measures. First, note that reader group differences emerged on the two verbal memory tests and on the visual–verbal association learning test administered in kindergarten, Rao's $F(6, 118) = 4.42, p < .01$, and that the AvIQNorm group performed statistically better ($p < .05$) than the VLG group on all three of these tests. The AvIQNorm group also performed statistically better than the VGG group ($p < .05$) on the word memory and visual–verbal learning test, but no better than the VGG group on the sentence memory test ($p > .05$). At the same time, the VGG group performed statistically better ($p < .05$) than the VLG group on the visual–verbal learning test, although not statistically better than the VLG group on the sentence memory and word memory tests.

On the verbal memory tests administered in first grade, the AvIQNorm group performed better than both the VLG and the VGG groups on the test evaluating serial recall of auditorily presented digits (WISC–R Digit Span), $F(2, 53) = 18.92, p < .01$, but the latter groups were not statistically different from each other ($p > .05$). The AvIQNorm group also performed statistically better than the VLG group on the tests (respectively) evaluating verbatim recall of sentences (TOLD–P:2 Sentence Imitation) and syntactic ordering of words held in working memory (Syntactic Word Order), Rao's $F(4, 124) = 7.83, p < .01$, and better ($p < .05$) than the VGG group on the Syntactic Word Order test. However, the AvIQNorm group performed no better ($p > .05$) than the VGG group on the Sentence Imitation test. Moreover, the VGG group performed statistically better than the VLG group on the Syntactic Word Order test, although no better than the VLG group on the Sentence Imitation test.

Reader group differences also emerged on the test evaluating phonological memory, $F(2, 63) = 4.02, p < .05$, which is, perhaps, the most direct measure of phonological coding ability included in the test battery. However, the only difference that was found to be statistically significant was that between the AvIQNorm group and the VLG group. In contrast, reader group differences, Rao's $F(8, 120) = 2.95, p < .05$, that emerged on tests that also depend heavily on phonological coding ability—that is, tests evaluating immediate and delayed recall of concrete and abstract words—favored the VGG group over the VLG group ($p < .05$) on three out of four of these tests and the AvIQNorm group over the VLG group ($p < .05$) on two out of four of the tests. Yet, there were no statistically significant differ-

Table 9
Verbal Memory and Verbal Learning Measures Administered to Normal Readers and Respective Tutored Groups in Kindergarten and First Grade

Measure	Normal readers: Average IQ (n = 28)	Normal readers: Above average IQ (n = 37)	Tutored groups: VLG (n = 19)	Tutored groups: LG (n = 18)	Tutored groups: GG (n = 18)	Tutored groups: VGG (n = 19)
		Kindergarten				
Sentence Memory[a]						
M	4.96	0.71	−0.88	−0.50	−0.61	−0.27
SD	1.04					
Word Memory[a,b]						
M	1.50	0.36	−1.03	−0.56	−0.77	−0.76
SD	0.79					
Visual–Verbal Learning[a,b,c]						
M	41.52	0.17	−2.13	−1.22	−1.13	−1.04
SD	4.44					
		First grade				
WISC–R Digit Span[a,b]						
M	8.22	0.45	−1.70	−0.93	−1.47	−1.65
SD	1.44					
TOLD–P:2 Sentence Imitation[a]						
M	18.89	0.61	−1.24	−0.55	−0.92	−0.72
SD	4.72					
Syntactic Word Order[a,b,c]						
M	20.04	0.09	−2.15	−1.71	−1.01	−1.00
SD	2.40					
Phonological Memory[a]						
M	14.50	0.52	−0.76	−0.50	−0.65	−0.22
SD	5.40					
Immediate Recall Concrete						
M	18.46	0.29	−0.03	−0.04	0.10	0.27
SD	5.46					
Immediate Recall Abstract[a,c]						
M	14.32	0.55	−0.61	−0.26	−0.17	0.17
SD	5.16					
Delayed Recall Concrete[c]						
M	3.04	0.10	−0.39	−0.11	0.31	0.45
SD	1.32					
Delayed Recall Abstract[a,c]						
M	1.96	0.01	−0.95	−0.80	−0.55	−0.25
SD	1.35					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the above average IQ normal reader group. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; WISC–R = Wechsler Intelligence Scale for Children—Revised; TOLD–P:2 = Test of Language Development—Primary:2; AvIQNorm = average IQ normal.
[a] Significant differences between the AvIQNorm group versus the VLG group. [b] Significant difference between the AvIQNorm group versus the VGG group. [c] Significant difference between the VGG group versus the VLG group.

ences ($p > .05$) between the AvIQNorm and the VGG groups on any of them.

These results are quite in keeping with the view that reading difficulties, in some children, are caused by phonological coding deficits that affect both working memory and reading subskills that depend, in part, on working memory, in particular, on name encoding, name retrieval, and visual–verbal association learning of the type entailed in learning to read.

Visual Processing

The finding of strong and consistent reader group differences on the language-based measures administered to our target children contrasts with the absence of strong and consistent reader group differences on the visual measures administered (Table 10). We should first remind the reader that there were no significant differences between the AvIQNorm readers and any of the tutored groups on the

Table 10
Visual-Processing Measures Administered to Normal Readers and Respective Tutored Groups in Kindergarten and First Grade

	Normal readers		Tutored groups			
Measure	Average IQ (n = 28)	Above average IQ (n = 37)	VLG (n = 19)	LG (n = 18)	GG (n = 18)	VGG (n = 19)
		Kindergarten				
WPPSI–R Block Design						
M	22.14	0.63	−0.41	−0.48	−0.76	−0.40
SD	5.37					
Visual Memory labelable						
M	1.52	0.46	−0.39	−0.56	−0.69	−0.45
SD	0.94					
Visual Memory nonlabelable						
M	5.59	0.28	−0.31	−0.41	−0.54	−0.55
SD	2.15					
		First grade				
WISC–R Block Design						
M	13.57	0.80	0.09	−0.30	−0.31	0.19
SD	7.16					
Visual Memory labelable[a,b]						
M	5.39	0.15	−1.38	−1.29	−0.81	−1.60
SD	1.17					
Visual Memory nonlabelable						
M	6.18	0.24	−0.41	−1.05	−0.18	−0.60
SD	1.87					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the above average IQ normal reader group. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth; WPPSI–R = Wechsler Preschool and Primary Scale of Intelligence—Revised; WISC–R = Wechsler Intelligence Scale for Children—Revised; AvIQNorm = average IQ normal.
[a] Significant difference between the AvIQNorm group versus the VLG group. [b] Significant difference between the AvIQNorm group versus the VGG group.

WISC–R Performance IQ, which is a composite derived from subtests that depend heavily on a variety of visual-processing and visuomotor abilities. Second, no significant differences emerged between any of the groups compared on the visual-processing tasks administered in kindergarten, specifically, WPPSI–R Block Design and Visual Memory for labelable and nonlabelable dot patterns. And, although the F ratio for the same tasks administered in first grade was statistically significant, Rao's $F(6, 122) = 3.96$, $p < .01$, the Newman-Keuls tests revealed that reader group differences were significant ($p < .05$) only on contrasts between the AvIQNorm and both the VLG and the VGG groups and only in the case of the visual memory task, which used dot patterns that could readily be labeled, for example, an array of dots that formed an uppercase *T*. It is not surprising to find differences between poor and normal readers on visual memory tasks when the visual stimuli can be readily labeled, and, in fact, this outcome has been obtained in previous research (Katz, Shankweiler, & Liberman, 1981).

Executive Functions

Finally, Table 11 presents results on the tests administered to evaluate attentional, organizational, and strategic planning abilities: a modified version of the Matching Familiar Figures Test (MFF) developed by Kagan (1965) and modified versions of the Target Search Test (TST) developed by Rudel et al. (1978). Analysis of results for kindergarten testing produced no significant differences between respective reader groups compared on these measures ($p > .05$). At second-grade testing, there was a significant group effect, Rao's $F(12, 110) = 2.75$, $p < .05$, although Newman-Keuls tests revealed that there were no group differences for the TST time variables. For the accuracy variables, a significant difference ($p < .05$) occurred only in the case of the letter cluster targets (e.g., pronounceable nonsense words). Both the AvIQNorm group and the VGG group performed better than the VLG group in detecting these targets, which, quite likely, reflects that children in the former groups were better able than children in the latter group to use verbal codes to assist in target detection. We,

Table 11
"Executive Function" Measures Administered to Normal Readers and Children in Respective Tutored Groups in Kindergarten and Second Grade

Measure	Normal readers: Average IQ (*n* = 28)	Normal readers: Above average IQ (*n* = 37)	Tutored groups: VLG (*n* = 19)	Tutored groups: LG (*n* = 18)	Tutored groups: GG (*n* = 18)	Tutored groups: VGG (*n* = 19)
		Kindergarten				
Matching Familiar Figures time						
M	5.99	0.26	−0.16	−0.22	−0.16	−0.14
SD	2.54					
Matching Familiar Figures error						
M	6.36	−0.04	0.83	0.54	0.82	0.57
SD	2.42					
Target Search time						
M	83.59	−0.23	0.35	0.25	0.11	0.32
SD	27.92					
Target Search correct						
M	16.18	−0.26	−0.45	−0.36	−1.18	−0.06
SD	2.04					
		Second grade				
Target Search Symbol time						
M	80.43	0.08	0.03	0.24	0.19	0.32
SD	32.76					
Target Search Number time						
M	148.45	0.20	0.86	0.71	0.55	0.66
SD	60.33					
Target Search Letter time						
M	121.30	−0.05	0.44	0.50	0.00	0.21
SD	52.48					
Target Search Symbol correct						
M	9.31	0.18	0.00	0.01	0.01	−0.13
SD	3.60					
Target Search Number correct						
M	11.62	0.18	−0.83	−0.57	0.05	0.08
SD	2.23					
Target Search Letter correct[a,b]						
M	12.54	0.04	−1.31	−0.59	−0.83	−0.06
SD	1.92					

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. Results are reported as effect sizes for the poor readers and for children in the above average IQ normal reader group. VLG = very limited growth; LG = limited growth; GG = good growth; VGG = very good growth.
[a] Significant difference between average IQ normal group versus the VLG group. [b] Significant difference between the VGG group versus the VLG group.

therefore, conclude that attentional, organizational, and strategic planning deficits were not a significant source of reading difficulty in most children in the population studied herein.

Analyses Involving All Tutored Groups

To summarize up to this point, the cognitive measures that reliably and strongly distinguished the normal readers from the worst achieving tutored group (VLG) and, quite frequently, the best achieving group (VGG) from the worst achieving tutored group, were those that evaluated phonologically based skills, such as phoneme segmentation, name encoding, name retrieval, and working memory. Such measures less reliably and less strongly distinguished the AvIQNorm readers from the best achieving tutored group (VGG), which was often found to be an intermediate group, having means between the normal readers and the worst achieving group, although differing from neither. In contrast, cognitive measures that evaluated semantic, syntactic, visual, and attentional abilities produced few significant differences between any of the groups compared. These results are consistent with the possibility that many, if not most, children in the VLG group were at the low end of the continuua of cognitive abilities underlying reading ability at the times they were tested, whereas many, if not most, children in the VGG group were substantially higher on these continuua, with children in the LG and GG groups being more often between children in the VLG and VGG groups. However, to evaluate the generality of observed effects, we performed another set of MANOVAs comparing the AvIQNorm group with reconstituted groups of tutored poor readers on the measures included in the kindergarten and first-grade test batteries. For these analyses, the tutored groups were reconstituted by splitting the entire sample (N = 74) at the median of the W slopes continuum, thereby combining the VLG group with the LG group (low reading growth [LRG]), and the GG group with the VGG group (high reading growth [HRG]). Effect sizes for the reconstituted tutored groups, relative to the AvIQNorm group, are presented in Table 12. Only those variables that yielded statistically significant differences between given groups appear in the table.

It can be seen that these analyses produced essentially the same pattern of results as the analyses involving the AvIQNorm, the VLG, and the VGG groups, insofar as the types of phonological and fluency measures that distinguished these respective groups in the previous analyses in many instances distinguished these groups in these analyses. In contrast, the semantic, syntactic, visual, and attentional measures did not reliably distinguish the latter groups, as in the previous analyses. The major effect of reconstituting the groups was to bring the HRG and LRG groups closer together on the cognitive measures that initially discriminated between the VGG and the VLG groups. This had the effect of reducing the number of statistically significant differences between the reconstituted tutored groups on these measures, while increasing the number of such differences between these (respective) groups and the AvIQNorm group. Nevertheless, the measures that most strongly distinguished the VLG and VGG groups in the previous analyses continued to distinguish the LRG and HRG groups in the additional analyses involving the latter groups, notably letter and number identification, rapid naming of letters and numbers, recall of concrete and abstract words, and processing sentences and words in working memory (respectively, Token Part V and Syntactic Word Order). Moreover, as we pointed out earlier, statistically significant differences emerged between the AvIQNorm group and the LRG group on the test of phoneme segmentation administered in kindergarten, with the HRG group being intermediate between the two, and differing from neither. However, differences between the AvIQNorm group versus the LRG and HRG groups (respectively) were found to be statistically significant on both first-grade winter and spring testing.

If the tasks that reliably discriminate between respective reader groups evaluate skills and abilities that are important for learning to read, then they should often be found to be significantly correlated with growth in (beginning) reading ability. Conversely, tasks that do not reliably discriminate between respective reader groups may evaluate skills and abilities that are less important for learning to read, and if this is the case, then such tasks should less often be found to be significantly correlated with growth in (beginning) reading ability. To ascertain the validity of these assertions, we computed simple correlations between each of the cog-

Table 12
Effect Sizes for High- and Low-Reading-Growth Children Relative to Average IQ Normal Readers on Kindergarten, First and Second Grade Skills and Abilities That Were Found to Differentiate AvIQNorm, VLG, and VGG Groups

Variable	Low reading growth (n = 37)	High reading growth (n = 37)
Letter identification K[a,b,c]	−2.56	−1.62
Word identification K[a,b]	−0.57	−0.59
Arithmetic (WPPSI–R) K[a,b]	−1.01	−0.77
Counting by 1s K[a,b]	−1.10	−0.71
Counting by 2s K[a,b]	−0.36	−0.39
Number Identification K[a,b,c]	−2.00	−1.13
WJ Applied Problems 1[a,b]	−0.70	−0.40
Phoneme Segmentation K[a]	−0.51	−0.34
Phoneme Segmentation 1W[a,b]	−1.24	−1.18
Phoneme Segmentation 1S[a,b]	−0.98	−0.91
RAN Objects time K[a,b]	0.77	0.60
RAN Objects errors K[a,b]	0.83	0.64
Rapid Articulation time K[a,b]	1.33	0.98
RAN Objects time 1[a,b]	0.67	0.57
RAN Letters time 1[a,c]	1.27	0.46
RAN Numbers time 1[a,c]	0.67	0.09
Boston Naming Test correct 1[a,b]	−1.08	−0.67
Token Test IV 1[a,b]	−1.39	−1.37
Token Test V 1[a,c]	−0.78	−0.19
Listening Comprehension[a,b]	−0.62	−0.81
Visual–Verbal Learning K[a,b]	−1.67	−1.08
Word Memory K[a,b]	−0.80	−0.77
Delayed Recall Concrete 1[c]	−0.25	0.38
Delayed Recall Abstract 1[a,c]	−0.88	−0.39
WISC–R Digit Span 1[a,b]	−1.38	−1.56
TOLD–P:2 Sentence Imitation 1[a,b]	−0.91	−0.82
Syntactic Word Order 1[a,b,c]	−1.93	−1.01
Phonological Memory[a]	−0.63	−0.43
Visual Memory Labelable 1[a,b]	−1.33	−1.22
Visual Memory Nonlabelable 1[a]	−0.72	−0.40
Target Search Numbers time 2[a,b]	0.79	0.61
Target Search Letters correct 2[a]	−0.96	−0.44
Target Search Numbers correct 2[a,c]	−0.70	0.07

Note. Tutored children are grouped by slopes for W scores obtained on the Basic Skills Cluster of the Woodcock Reading Mastery Tests—Revised from kindergarten through fall of second grade. K = kindergarten; 1 = first grade; 2 = second grade; W = winter; S = spring; WPPSI–R = Wechsler Preschool and Primary Scale of Intelligence—Revised; WJ = Woodcock-Johnson Psycho-Educational Battery; RAN = Rapid Automatized Naming; WISC–R = Wechsler Intelligence Scale for Children—Revised; AvIQNorm = average IQ normal; VLG = very low growth; VGG = very good growth. TOLD–P:2 = Test of Language Development—Primary:2.
[a] Significant difference between the AvIQNorm versus low reading growth group. [b] Significant difference between the AvIQNorm versus high reading growth group. [c] Significant difference between the high and low reading growth groups.

nitive tasks given at kindergarten, first-, and second-grade testing and growth in reading subskills, as defined by individual slopes derived from the raw scores on the Word Identification and Word Attack subtests of the WRMT–R, respectively. We computed separate correlations for children in the tutored groups combined ($n = 74$) and for children in the normal reader groups combined ($n = 65$). Table 13 presents these correlations. The table lists only those variables that produced statistically significant correlations, when either the slope for the WRMT–R Word Identification subtest or the slope for the WRMT–R Word Attack subtest was the correlated measure.

It can be seen that for both the tutored children and the normal readers, it is essentially the same assortment of variables that reliably discriminated between the various reader groups compared in the analyses discussed thus far that are significantly correlated with the slopes derived from either the Word Identification or the Word Attack subtests. In contrast, none of the semantic or visual measures was found to be significantly correlated with either of the slope measures. The only syntactic measures that were significantly correlated with either of the slope measures were Parts IV and V of the Token Test, which, as we asserted earlier, make heavy demands on working memory. Also, although certain of the target search tasks that were used to evaluate attentional and strategic planning abilities produced significant correlations with the growth measures, it was the tasks involving detection of letters and numbers that produced these correlations. As we indicated earlier, performance on both of these tasks could be influenced by verbal coding ability, and it may be individual differences in verbal coding ability that account for observed correlations rather than individual differences in attentional and strategic planning abilities.

Table 13
Pearson Correlations That Were Found to Be Statistically Significant Between Kindergarten, First, and Second-Grade Cognitive Measures and Slopes Representing Growth on Either the Word Identification or Word Attack Subtests of the Woodcock Reading Mastery Tests—Revised

Variable	Word Identification slope	Word Attack slope
Panel A: Tutored children ($n = 74$)[a]		
Letter Identification K	.39	.29
Counting by 1s K	.27	.29
Number Identification K	.34	.22
Phoneme Segmentation K	.19	.28
Phoneme Segmentation 1s	.18	.29
WJ applied Problems 1	.23	.20
RAN Letters Time 1	−.49	−.36
RAN Numbers Time 1	−.41	−.27
Token Test V1	.27	.25
Listening Comprehension 1	−.10	−.26
Syntactic Word Order 1	.31	.15
Delayed Recall Concrete 1	.32	.24
Immediate Recall Abstract 1	.23	.33
Delayed Recall Abstract 1	.30	.24
Target Search Numbers Correct 2	.25	.13
Target Search Letters Correct 2	.29	.25
Panel B: Normal Readers ($n = 65$)[b]		
Print Awareness K	0.28	0.30
Phoneme Segmentation 1W	0.38	0.43
Phoneme Segmentation 1S	0.27	0.32
RAN Objects Time K	−0.28	−0.29
RAN Letters Time 1	−0.40	−0.34
RAN Numbers Time 1	−0.38	−0.36
Token Test IV 1	0.25	0.18
Token Test V 1	0.32	0.29
Listening Comprehension 1	0.28	0.17
Syntactic Word Order 1	0.29	0.43
Phonological Memory 1	0.25	0.29
Target Search Numbers Correct 2	0.27	0.33
Target Search Numbers Time 2	0.25	0.22

Note. K = kindergarten; 1 = first grade; 2 = second grade; S = spring; W = winter. WJ = Woodcock-Johnson Psycho-Educational Battery; RAN = Rapid Automatized Naming.
[a] For the tutored children, correlations between $r = .22$ and $r = .27$ are significant at $p < .05$. Correlations larger than $r = .27$ are significant at $p < .01$, for two-tailed tests. [b] For normal readers, correlations between .25 and .30 are significant at $p < .05$. Correlations greater than .30 are significant at $p < .01$, for two-tailed tests.

Comparisons of Average Versus Above Average Normal Readers

Finally, given that the AvIQNorm and the AbAvIQNorm readers did not differ on any of the reading measures, we conducted significance tests to evaluate differences between these two groups on all of the measures included in the kindergarten and first-grade cognitive batteries. It will suffice to point out that although there were several statistically significant differences between these groups on experience-based and cognitive tasks such as those included on the intelligence measures (e.g., general knowledge, vocabulary, visuospatial abilities, etc.), there were very few significant differences between these groups on measures of phonologically based skills such as phoneme segmentation, naming, verbal memory, and visual–verbal learning. The few that achieved statistical significance ($p < .05$) in favor of above average IQ children, were sentence memory (kindergarten), oral cloze (first grade), immediate recall of abstract words (first grade), TOLD–P:2 Sentence Imitation (first grade), and Listening Comprehension (first grade).

GENERAL DISCUSSION

In discussing a view articulated by Clay (1987), we found merit in her suggestion that the major impediment to distinguishing between reading difficulties caused primarily by basic cognitive deficits and reading difficulties caused primarily by experiential and instructional deficits is the failure to control for the child's educational history. We pointed out, in accord with Clay's concerns, that although psychological, genetic, neuropathological, and electrophysiological studies provide highly suggestive evidence that reading disability in some poor readers may be caused by language-based deficits of constitutional origin, especially phonological deficits, results from those studies are open to question.

This is because none controlled for inadequate instruction and other experiential factors that could be mistaken for basic cognitive deficits, in terms of their effects on skills and abilities that underlie reading. We also pointed out that previous intervention studies conducted by Clay (1985) and others (Iversen & Tunmer, 1993; Pinnell, 1989; Wasik & Slavin, 1993) have quite convincingly demonstrated that most reading-impaired children can acquire at least grade-level reading skills if they receive early and labor-intensive remediation to correct their deficiencies. This, of course, suggests that the majority of children who might be diagnosed as "reading disabled" are impaired by experiential and instructional deficits rather than basic cognitive deficits. However, the results of these studies were, themselves, inconclusive, because none compared the cognitive abilities of children who were readily remediated with the cognitive abilities of children who were difficult to remediate, to ascertain whether cognitive deficiencies believed to be causally related to reading disability are found more often in the latter group than in the former. Moreover, none attempted to evaluate the prereading skills and abilities of the children in these two groups before they received formal instruction in reading.

To accomplish these objectives, we conducted a longitudinal study that incorporated response to early intervention as the initial diagnostic indicant, along with assessment of prereading skills and cognitive abilities in children who were identified as "disabled readers," and the results are salutary. In accord with the early intervention studies conducted previously, we found that the largest percentage (67.1%) of poor readers who received daily one-to-one tutoring scored within the average or above average ranges on standardized tests of reading achievement after only one school semester of tutoring. Conversely, a much smaller percentage (32.9%) of these children scored below the average range and the smallest percentage (15%) scored in the severely impaired range (<15%). To put these figures in terms of percentages based on the total number of children in the population from which the tutored children were drawn (n = 827),[9] 12 children, or 1.5% of this population, scored below the 15th percentile on a composite measure of reading ability after one semester of remediation, whereas 25 children, or 3%, scored below the 30th percentile after one semester of remediation. These figures represent a substantial reduction over the 9% figure (118 out of 1,284 = .09) that emerges as the estimate of the percentage of reading-disabled children in the population when one uses only the exclusionary criteria that are typically used to identify such children. Moreover, when we grouped the tutored children on the basis of growth in reading ability during the period preceding and immediately following intervention (operationally defined by individual slopes on a composite measure of reading ability), we found that the groups were rank ordered such that the group that manifested the most accelerated rate of growth in reading subskills, as a function of our intervention program, approached the level of the normal readers and maintained their advantage over children in all other groups thereafter. At the same time, children in the group that manifested the least accelerated rate of growth in these subskills performed well below the children in all other groups thereafter, with each respective group maintaining its relative status throughout the period evaluated.

These results, like those obtained in the previous intervention studies, are consistent with Clay's (1987) contention that most impaired readers, who might be classified as learning disabled, are probably not learning disabled in the stereotypical sense in which this term is used, that is, as a label for someone whose learning difficulties are presumed to be of constitutional origin. However, our results are also consistent with the possibility that the learning difficulties of at least some of these children may well be of constitutional origin and that they are, quite likely, associated with a less than adequate mix of cognitive abilities underlying reading ability, especially phonological abilities. There are several layers of additional evidence that support both possibilities.

First, in view of the finding that most of the tutored children became at least average-level readers in two, if not one, semesters of remediation, it seems reasonable to hold with our earlier assertion that the sole use of exclusionary criteria to identify "disabled readers" does not guarantee that children so identified are truly "disabled," even if they are severely impaired readers. Yet, our use of these criteria did have the apparent effect of excluding from the sample children whose reading problems could have been attributed to general learning difficulties or pervasive knowledge deficits that could accrue because of lack of environmental stimulation. That this was the case is supported by our finding that children in each of the tutored groups performed as well as the AvIQNorm readers on most of the tests of reading readiness administered in kindergarten, specifically those evaluating print awareness, concepts of print, general knowledge, and general concept development. Additional support is provided by the absence of differences between these respective groups and the AvIQNorm readers on both the Verbal and Performance subtests of the WISC–R administered in first grade, and by the fact that children in the tutored groups scored at least within the average range on measures of math achievement administered in first grade.

However, when the normal readers were compared with children in the tutored groups on measures evaluating cognitive abilities presumed to underlie reading ability, the

[9] To calculate the percentage of tutored children who continued to fall within the "disabled" categories following intervention, we used as the population base the total number (N = 827) yielded by multiplying the number of children in the total population of available children (N = 1,284) by the percentage corresponding to the total number of identified poor readers who received one-to-one tutoring (76 out of 118 = 64%; 1,284 × 64% = 827). Accordingly, the number of tutored children who scored below the 15th percentile on the BSC after one semester of remediation represents 1.5% (12 out of 827) of the population from which these children were drawn, whereas the number of tutored children who scored below the 30th percentile on the BSC after the same period represents 3% (25 out of 827) of the population from which they were drawn.

groups were found to differ statistically on some, but not all, of these measures, and the results have implications for etiological hypotheses that were in need of further evaluation. As we have already indicated, this study provides more definitive documentation than do previous studies (e.g., Vellutino & Scanlon, 1987a; Vellutino et al., 1988; Vellutino et al., 1994; Vellutino et al., 1995) that reading difficulties in beginning readers from the population sampled here (middle- to upper middle-class children) are probably not caused by either semantic deficits or visual-processing deficits of the types most often proposed, given that no strong or reliable differences emerged between and among respective groups on measures of these abilities. This pattern of results was observed both on kindergarten testing, before children were identified as poor or normal readers, and on first-grade testing, after they were so identified. Similarly, neither kindergarten nor first-grade testing produced reliable differences between and among these groups on the various measures evaluating syntactic competence, except for those that place a heavy load on working memory. These results bring to the fore perhaps the most important collection of findings that emerged from this study vis-à-vis the major purpose for conducting the study.

We indicated earlier that the most reliable and convergent findings from studies evaluating the etiology of reading disability implicate deficiencies in phonologically based skills as basic causes of the disorder. The present findings add considerable weight to results obtained in these studies, not only because previously observed differences between poor and normal readers on tasks evaluating such skills were replicated in preliterate children who were subsequently identified as poor versus normal readers, but because performance on these tasks also differentiated tutored children who were difficult to remediate and tutored children who were readily remediated. For example, in accord with results obtained in previous research, the normal readers generally performed better than children in the VLG and VGG groups on the tests of phoneme segmentation administered in kindergarten and on those administered in winter and spring of first grade. However, although group differences were found to be statistically significant only on winter and spring testing, the normal readers performed significantly better than the two worst achieving tutored groups combined (VLG and LG) at kindergarten testing and better than both this group and the two best achieving tutored groups combined (GG and VGG) at first-grade winter and spring testing. The normal readers also performed statistically better than children in the tutored groups on the test of phonetic decoding on all occasions on which this test was administered, both before and after remedial intervention. Yet, it is of some significance that, following only one semester of intervention, the children in the VGG group performed statistically better than the children in the VLG group on the tests evaluating phoneme segmentation and phonetic decoding ability and approximated the level of the normal readers on these tests. Because the intervention program included activities designed to improve the child's facility in phoneme segmentation and phonetic decoding, as well as his or her facility in word identification and text processing, we interpret these results to mean that acquiring each of these skills was inherently more difficult for children in the worst achieving groups than for children in the better achieving groups. This, we suggest, is because children in the former groups were more often afflicted by phonological coding deficits than were children in the latter groups. (However, the relationship was not strictly linear in the case of phoneme segmentation in that children in the LG group performed as well as or better than children in the GG and VGG groups on first-grade testing, suggesting that the phonological abilities of some children in the LG group were as good as or better than those of some children in the GG and VGG groups.)[10]

Additional support for the inferred relationship between phonological coding ability and the ability to profit from reading instruction is provided by the emergence of reader group differences on tests evaluating phonological skills such as rapid naming, confrontational naming, visual–verbal learning, and verbal memory. Many of these tasks distinguished between the best and worst achieving tutored groups, and all of them distinguished between the worst achieving tutored group and the normal reader group(s). However, these tasks did not always distinguish between the normal readers and the best achieving tutored group. These findings are complemented by the finding that most of the phonological measures were found to be correlated with individual slopes representing growth in reading subskills in both the tutored children and in the normal readers. It therefore seems reasonable to infer that deficiencies in both reading and phonologically based skills that have been found to be intrinsically and, in some instances, causally related to deficiencies in reading (e.g., Blachman, 1994; Bradley & Bryant, 1983; Denckla & Rudel, 1976a, 1976b; Lundberg et al., 1988; Vellutino & Scanlon, 1987a, 1987b; Vellutino et al., 1995; Wolf, 1984) are, in many cases, due primarily to basic deficits in phonological coding ability, perhaps associated with inherent differences in this ability.

The inference that reader group differences on tasks evaluating phonologically based abilities such as phoneme segmentation, phonetic decoding, naming, and verbal memory may, in many cases, be due in part to inherent deficits rather than to deficits that could accrue as a consequence of reading difficulties (Bryant & Goswami, 1986) is supported

[10] The possibility that the phonological abilities of some children in the LG group were as good as or better than the phonological abilities of some children in the GG and VGG groups is also suggested in the finding that on a few measures evaluating phonologically based skills, other than phoneme segmentation, performance in the LG group (in terms of mean effect sizes) was comparable to or better than performance in the GG and VGG groups, for example, rapid naming of colors, word memory, memory for digits, and memory for sentences. This pattern of results suggests that the limited growth in reading characteristic of children in the LG group was attributable, in some cases, to factors other than phonological coding deficits. It also suggests that the probability of finding children whose reading difficulties are caused primarily by cognitive deficits would be greater in the case of those in the VLG group than in the case of those in any of the other groups, including the LG group.

by our finding of reader group differences on such tasks at the beginning stages of literacy development before Matthew effects (Stanovich, 1986) could have begun to accumulate. Especially supportive are the differences that emerged between respective reader groups on the cognitive tasks given the children in these groups when they were in kindergarten, before most of them had begun to learn to read.

However, this inference carries with it certain disclaimers. First, to conclude from our data that reading difficulties in beginning readers may often be associated with phonological coding deficits is, quite simply, to acknowledge the heavy dependence of reading in an alphabetically based orthography on the phonology of language (Vellutino, 1991). It seems clear that phonological coding ability is the primary determinant of the child's success in mastering the alphabetic code and in learning to attach names to printed words as wholes and that he or she must acquire facility in both subskills to learn to read. Yet these assertions in no way deny the importance of the semantic, syntactic, and visual components of written language as partial determinants of the child's ability to learn to read. Knowledge of word meanings and knowledge of syntax contribute, directly or indirectly, to the acquisition of facility in word identification, and both are critically important for reading comprehension. It follows that deficiencies in one or both of these areas could be a source of difficulty in either word identification or comprehension. However, one would not expect to find a high incidence of semantic or syntactic deficits in beginning readers from middle- to upper middle-class populations, and this is, no doubt, one reason that we found no reliable differences between poor and normal readers or between any of the tutored groups, on semantic and syntactic measures. Another reason is that the semantic and syntactic attributes of written English may carry less weight as determinants of success in beginning reading than do its phonological attributes, given that English orthography is derived from an alphabet. There is, in fact, independent evidence supporting this possibility, as we indicated earlier (e.g., Vellutino et al., 1991; Vellutino et al., 1994).

Similarly, the beginning reader must distinguish the graphic and orthographic attributes of printed words in learning to identify them, but because of the formidable load on visual memory occasioned by the alphabetic properties of written English, reliable identification of these words will ultimately require that he or she stores quality representations of their structural redundancies. Indeed, it is in the process of capitalizing on these redundancies that the child increasingly discovers the features that distinguish between words that have a high degree of graphic and orthographic similarity (Gough & Hillinger, 1980; Gough & Tunmer, 1986). Yet, the success of each of these enterprises in this writing system is determined primarily by the child's ability to master the alphabetic code, and this ability depends, in turn, on phonological coding ability. This, of course, implies that visual abilities are less important determinants of success in beginning reading than are phonological abilities, and there is independent evidence to support this suggestion as well (Vellutino et al., 1991; Vellutino et al., 1994). Thus, to find, as we did, that the best and worst achieving tutored groups, as well as the worst achieving and normal reader groups, differed statistically on measures of phonological skills, but not on measures of semantic, syntactic, and visual skills, is logically consistent and coherent.

As a final disclaimer, we wish to point out that although the present findings are consistent with the possibility that reading difficulties in some children may be caused by phonological deficiencies of constitutional origin, we do not believe that such deficiencies necessarily emanate from neurological damage. However, this possibility is not ruled out in any given case. As we suggested earlier, both reading ability and the cognitive abilities underlying reading ability might better be placed on continuua, such that those who have an optimal mix of these abilities (e.g., strong phonological skills) will tend to have the greatest amount of success in beginning reading, whereas those who have a less optimal mix of these abilities (e.g., weak phonological skills) will tend to have the least success in beginning reading, all other factors being equal. Results on all of the reading measures administered provide the strongest support for this suggestion, insofar as the groups that performed at the low end of the continuua on tests evaluating word identification, oral reading, and reading comprehension—that is, the two worst achieving groups (VLG and LG)—also performed at the low end of the continuua on tests evaluating phonetic decoding ability, which is a critically important phonological skill underlying beginning reading. The converse was true in the case of the normal readers and the two best achieving groups (GG and VGG), and it is of some significance that performance in these groups on tests evaluating phoneme segmentation, naming, verbal memory, and other tests that rely heavily on phonological coding abilities often approximated a linear trend across respective reader groups. This pattern was especially evident in contrasts involving the reconstituted tutored groups (VLG and LG vs. GG and VGG) relative to the normal readers (see Table 12), and the data in general speak for the validity of a continuum-based conceptualization of reading ability and disability (see Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992, for additional evidence in support of this conclusion).

However, we should remind the reader, in connection with these assertions, that children in the tutored groups generally performed below the normal readers (as evidenced, in most instances, by negative effect sizes), not only on measures of phonologically based skills, but also on measures of semantic, syntactic, and visual abilities that might contribute directly or indirectly to reading achievement and response to remediation. Although reader group differences in these latter abilities were rarely found to be statistically significant, it must nevertheless be acknowledged that the tendency to fall below the normal readers was a general tendency even in the case of the best achieving tutored group (VGG), and this finding could be taken as evidence that reading-related cognitive abilities in the tutored children were generally subnormal. We think that this interpretation is probably incorrect and suggest that there are two more plausible interpretations of this pattern of results. One is that the AvIQNorm group (the group on which effect sizes were based) generally scored above national norms on the standardized reading test (e.g., 72nd percentile on the BSC of the WRMT–R), which suggests that they were generally above average rather than average readers. Thus, it should not be surprising to find that they typically performed above the level of the tutored children on measures of cognitive abilities underlying reading ability, including those abilities that carry somewhat less weight as determinants of success in beginning reading. Indeed, this pattern of results suggests that the normal readers were characterized by a more optimal mix of reading-related cognitive abilities than were the tutored children, even those in the best achieving group. Yet, the cognitive abilities of the better achievers were, nevertheless, within normal limits in many cases, which, of course, justifies our suggestion that most of these children were not "disabled" readers in the sense in which this term is typically used. In contrast, the cognitive abilities of the worst achievers, in many more cases, were outside normal limits, which justifies our suggestion that at least some of these children may have been truly "disabled" readers.

A second reason for doubting the cognitive deficit explanation of negative effect sizes in the better achievers is that performance on some of the cognitive measures could have been adversely affected by limitations in the child's experiences before school entry in roughly the same manner as rudimentary literacy skills could have been adversely affected by limitations in such experiences. Of course, these explanations are not mutually exclusive, and they are both consistent with a continuum-based view of reading ability and disability.

The foregoing disclaimers notwithstanding, our findings also have important implications for several other issues of concern to both researchers and practitioners. First, it should be apparent, both from this study and from other intervention studies that have appeared in the literature (Clay, 1985; Iversen & Tunmer, 1993; Pinnell, 1989; Wasik & Slavin, 1993), that to render a diagnosis of specific reading disability in the absence of early and labor-intensive remedial reading that has been tailored to the child's individual needs is, at best, a hazardous and dubious enterprise, given all of the stereotypes attached to this diagnosis. Our findings add to the findings of these earlier studies by demonstrating that one can increase the probability of validating the diagnosis if one combines impressions and outcomes derived from early, labor-intensive, and individualized remediation with results of relevant psychological and educational testing in evaluating the etiology of a child's difficulties in learning to read.

A second important implication of the present findings is concerned with early identification of children at risk for reading difficulties. Given that the impaired readers who received daily tutoring were found to perform below the normal readers not only on tests of reading achievement and phonological abilities administered in first grade, but also on tests of rudimentary literacy skills and phonological abilities administered in kindergarten, there is reason to be sanguine about the possibility that children at risk for reading difficulties can be identified and treated even before they are exposed to formal instruction in reading. By attempting to remediate deficiencies in children who are lacking in rudimentary reading skills, we may not only increase the probability of providing them with the foundational skills necessary for success in beginning reading, but by virtue of the knowledge we may acquire about their ability to profit from such remediation, we may be better able to tailor subsequent instruction to their individual needs. Such feedback may also prove to be helpful in validating and qualifying results of kindergarten screening in a way that would increasingly improve our ability to identify those children who may be at risk for reading difficulties. We have discussed, elsewhere, results of a prediction study that emerged from the current project, which provides additional confirmation for these assertions (Scanlon et al., 1995).

Finally, our data question the utility and widespread use of IQ–achievement discrepancy definitions of reading disability based on commonly used tests of intelligence such as the WISC–R (Rutter & Yule, 1975). We say this not only because the AvIQNorm and AbAvIQNorm readers in our study were found to perform at comparable levels on virtually all of the reading measures administered, but also because these two groups were not found to differ substantially or reliably on measures of phonologically based skills and abilities similar to those that have been consistently found to distinguish between poor and normal readers. At the same time, children in the worst achieving tutored groups were found to perform statistically below both the AvIQNorm readers and children in the best achieving tutored groups on these measures, despite the fact that they were found not to differ on the intelligence measures. Similar results have been obtained in studies recently reported by Siegel (1988), Fletcher et al. (1994), and Stanovich and Siegel (1994). The combined data sets strongly suggest that there is not the kind of linear relationship between IQ and reading ability assumed by IQ–achievement discrepancy definitions of reading disability, such that an average IQ implies average reading ability, a high average IQ implies high average reading ability, and a very high IQ implies superior reading ability, at least not when reading is simply defined as the ability to learn to "decode" printed words. The data also strongly suggest that many of the skills and abilities evaluated by intelligence tests such as the WISC–R are not as important for success in beginning reading as are phonological skills such as phoneme segmentation, phonetic decoding, and name encoding and retrieval. In fact, a reasonable interpretation of the present findings is that one needs little more than average intelligence to learn to decode print and that, given at least this level of intellectual ability, degree of facility in print decoding will ultimately be determined by degree of facility in phonological skills such as phonetic decoding, name encoding, and name retrieval. If these interpretations prove to be correct, then IQ–achievement discrepancy definitions of reading disability would be invalidated, and their widespread use in research, education, and clinical practice would need to be reconsidered. (See also Siegel, 1989, for additional evidence and an excellent discussion of these issues.)

In sum, this study is the first study that we know of to combine longitudinal analysis and response to early and labor-intensive intervention with cognitive profile analysis as the primary vehicles for evaluating the etiology of reading disability. Our results suggest that, although reading difficulties in most children from middle- to upper middle-class backgrounds are quite likely caused by experiential and instructional deficits, there are substantial numbers of these children whose reading difficulties may be caused by basic phonological coding deficits that may well be of constitutional origin. The data, in effect, validate the highly convergent findings from previous research implicating phonological coding deficits as a probable cause of reading disability in such children.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Bakker, D. J. (1972). *Temporal order in disturbed reading—Developmental and neuropsychological aspects in normal and reading-retarded children.* Rotterdam, The Netherlands: Rotterdam University Press.

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26,* 49–66.

Benton, A. L., & Hamsher, A. (1989). *The controlled word association test.* Iowa City, IA: AJA Associates.

Birch, H. G. (1962). Dyslexia and maturation of visual function. In J. Money (Ed.), *Reading disability: Progress and research needs in dyslexia* (pp. 161–169). Baltimore: Johns Hopkins Press.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76,* 610–622.

Blachman, B. A. (1994). Early literacy acquisition—The role of phonological awareness. In G. Wallach & K. Butler (Eds.), *Language learning disabilities in school-age children and adolescents: Some underlying principles and applications* (pp. 253–274). Columbus, OH: Merrill.

Blachman, B. A., Ball, E. W., Black, R. S., & Tangel, D. M. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms. *Reading and Writing: An International Journal, 6,* 1–18.

Bock, R. D. (1983). The discrete Bayesian. In H. Wainer & S. Messick (Eds.), *Principles of modern psychological measurement.* Hillsdale, NJ: Erlbaum.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 303,* 419–421.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35,* 345–367.

Breitmeyer, B. G. (1989). A visual based deficit in specific reading disability. *Irish Journal of Psychology, 10,* 534–541.

Brewer, W. F. (1967). *Paired-associate learning of dyslexic children.* Unpublished doctoral dissertation, University of Iowa.

Brittain, M. D. (1970). Inflectional performance and early reading achievement. *Reading Research Quarterly, 6,* 34–48.

Bruck, M. (1990). Word recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology, 26,* 439–454.

Bruck, M. (1992). The persistence of dyslexics' phonological awareness skills. *Developmental Psychology, 28,* 874–886.

Bryan, T., Donahue, M., & Pearl, R. (1981). Learning disabled children's peer interactions during a small group problem-solving task. *Learning Disability Quarterly, 4,* 13–22.

Bryant, P. E., & Goswami, U. (1986). Strengths and weaknesses of the reading level design: A comment on Backman, Mamen, and Ferguson. *Psychological Bulletin, 100,* 101–103.

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

Buschke, H. (1973). Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior, 12,* 543–550.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2,* 201–212.

Byrne, B., & Fielding-Barnsley, R. (1990). Acquiring the alphabetic principle: A case for teaching recognition of phoneme identity. *Journal of Educational Psychology, 82,* 805–812.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83,* 451–455.

Cardon, L. R., Smith, S. D., Fulker, D. W., Kimberling, W. J., Pennington, B. F., & DeFries, J. C. (1994, October 14). Quantitative trait locus for reading disability on chromosome 6. *Science, 226,* 276–279.

Catts, H. W. (1986). Speech production/phonological deficits in reading disordered children. *Journal of Learning Disabilities, 19,* 504–508.

Clay, M. M. (1985). *The early detection of reading difficulties* (3rd ed.) Auckland, New Zealand: Heinemann.

Clay, M. M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies, 22,* 155–173.

Decker, S. N., & Vandenberg, S. G. (1985). Colorado twin study of reading disability. In D. B. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 123–135). Parkton, MD: York Press.

DeFries, J. C. (1985). Colorado reading project. In D. B. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 107–122). Parkton, MD: York Press.

DeFries, J. C., Fulker, D. W., & LaBuda, M. C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature, 329,* 537–539.

Denckla, M. B., & Rudel, R. (1976a). Naming of pictured objects by dyslexic and other learning disabled children. *Brain and Language, 39,* 1–15.

Denckla, M. B., & Rudel, R. (1976b). Rapid "automatized" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471–479.

DiSimoni, F. (1978). *Token test for children.* Allen, TX: DLM Teaching Resources.

Donahue, M. (1986). Linguistic and communicative development in learning disabled children. In S. J. Ceci (Ed.), *Handbook of cognitive, social and neuropsychological aspects of learning disabilities* (Vol. 1, pp. 263–289). Hillsdale, NJ: Erlbaum.

Douglas, V. I. (1972). Stop, look and listen: The problem of sustained attention and impulse control in hyperactive and normal children. *Canadian Journal of Behavioral Science, 4,* 259–281.

Duffy, F. H., Denckla, M. B., Bartels, R. H., & Sandini, G. (1980). Dyslexia: Regional differences in brain electrical activity by topographic mapping. *Annals of Neurology, 7,* 412–420.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Dykman, R. A., Ackerman, P. T., & Holcomb, P. J. (1985). Reading disabled and ADD children: Similarities and differences. In D. B. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 47–62). Parkton, MD: York Press.

Ehri, L. (1989). The development of spelling knowledge and its role in reading acquisition and reading disability. *Journal of Learning Disabilities, 22,* 356–365.

Fildes, L. G. (1921). A psychological inquiry into the nature of the condition known as congenital word-blindness. *Brain, 44,* 286–307.

Filipek, P. A. (1995). Neurobiologic correlates of developmental dyslexia: How do dyslexics' brains differ from those of normal readers? *Journal of Child Neurology, 10,* 862–869.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Steubing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 6–23.

Flood, J., & Menyuk, P. (1983). The development of metalinguistic awareness and its relation to reading achievement. *Journal of Applied Developmental Psychology, 4,* 65–80.

Foorman, B. R., Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology, 83,* 456–469.

Fowler, A. E. (1988). Grammaticality judgment and reading skill in Grade 2. *Annals of Dyslexia, 38,* 73–94.

Fox, B., & Routh, D. K. (1980). Phonemic analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9,* 115–119.

Fry, M. A., Johnson, C. S., & Muehl, S. (1970). Oral language production in relation to reading achievement among select second graders. In D. J. Bakker & P. Satz (Eds.), *Specific reading disability: Advances in theory and method* (pp. 123–146). Rotterdam, The Netherlands: Rotterdam University Press.

Galaburda, A. (1983). Developmental dyslexia: Current anatomical research. *Annals of Dyslexia, 33,* 41–53.

Galaburda, A. M., & Kemper, T. L. (1979). Cytoarchitectonic abnormalities in developmental dyslexia: A case study. *Annals of Neurology, 6,* 94–100.

Gascon, G., & Goodglass, H. (1970). Reading retardation and the information content of stimuli in paired associate learning. *Cortex, 6,* 417–429.

Getman, G. N. (1985). A commentary on vision. *Journal of Learning Disabilities, 18,* 505–511.

Goldman, S. R. (1976). Reading skill and the minimum distance principle: A comparison of listening and reading comprehension. *Journal of Experimental Child Psychology, 22,* 123–142.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society, 30,* 179–196.

Gough, P. B., & Tunmer, 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.

Harter, M. R., Anllo-Vento, L., Wood, F. B., & Schroeder, M. M. (1988). Separate brain potential characteristics in children with reading disability and attention deficit disorder: Color and letter relevance effects. *Brain and Cognition, 7,* 115–140.

Hermann, K. (1959). *Reading disability.* Copenhagen, Denmark: Munksgaard.

Huba, M. E., & Kontos, S. (1985). Measuring print awareness in young children. *Journal of Educational Research, 78,* 272–279.

Hulme, C. (1988). The implausibility of low-level visual deficits as a cause of children's reading difficulties. *Cognitive Neuropsychology, 5,* 369–374.

Hynd, G. W., & Semrud-Clikeman, M. (1989). Dyslexia and brain morphology. *Psychological Bulletin, 106,* 447–482.

Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the reading recovery program. *Journal of Educational Psychology, 85,* 112–126.

Kagan, T. (1965). Reflectivity–impulsivity and reading ability in primary grade children. *Child Development, 36,* 609–628.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test.* Philadelphia: Lea & Febiger.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32,* 474–484.

Kavale, K. (1982). A comparison of learning disabled and normal children on the Boehm test of basic concepts. *Journal of Learning Disabilities, 15,* 160–161.

Kinsbourne, M., & Warrington, E. K. (1963). Developmental factors in reading and writing backwardness. *British Journal of Psychology, 54,* 145–156.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201–212.

Loban, W. (1963). *The language of elementary school children* (NCTE Research Report No. 1). Urbana, IL: National Council of Teachers of English.

Lovegrove, W., Martin, F., & Slaghuis, W. (1986). A theoretical and experimental case for a visual deficit in specific reading disability. *Cognitive Neuropsychology, 3,* 225–267.

Lundberg, I., Frost, J., & Petersen, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263–285.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21,* 159–173.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word storage in relation to reading ability. *Memory and Cognition, 8,* 329–335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11,* 627–643.

Martin, F., & Lovegrove, W. (1984). The effects of field size and luminance on contrast sensitivity differences between specifically reading disabled and normal children. *Neuropsychologia, 22,* 73–77.

Morrison, F. J., & Manis, F. R. (1982). Cognitive processes and reading disability: A critique and proposal. In C. J. Brainerd & M. Pressley (Eds.), *Verbal processes in children* (pp. 59–93). New York: Springer-Verlag.

Newcomer, P. L., & Hammill, D. D. (1991). *Test of Language Development—Primary: 2.* Austin, TX: Pro-Ed.

Olson, R. K., Kliegl, R., & Davidson, B. J. (1983). Dyslexic and normal readers' eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 816–825.

Olson, R., Wise, B., Conners, F., & Rack, J. (1990). Organization, heritability, and remediation of component word recognition and language skills in disabled readers. In T. H. Carr & B. A. Levy

(Eds.), *Reading and its development: Component skills approaches* (pp. 261–322). New York: Academic Press.

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

Orton, S. T. (1925). "Word-blindness" in school children. *Archives of Neurology and Psychiatry, 14,* 581–615.

Pavlidis, G. T. (1981). Do eye movements hold the key to dyslexia? *Neuropsychologia, 19,* 57–64.

Perfetti, C. A., Beck, I., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283–319.

Piaget, J. (1952). *The origins of intelligence.* New York: International Universities Press.

Pinnell, G. S. (1989). Reading recovery: Helping at-risk children learn to read. *Elementary School Journal, 90,* 161–184.

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

Rudel, R. G., Denckla, M. B., & Broman, M. (1978). Rapid silent response to repeated target symbols by dyslexic and nondyslexic children. *Brain and Language, 6,* 52–62.

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

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. 313–348). New York: Oxford University Press.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first grade reading: Selected results from a longitudinal study. *Mental Retardation and Development Disabilities, 2,* 54–63.

Scanlon, D. M., Vellutino, F. R., Small, S. G., Chen, R., & Denckla, M. B. (1995). *Prediction of "reading disability" and progress in remediation using a kindergarten test battery.* Unpublished manuscript.

Seaman, M. A., Levin, J. R., & Serlin, R. C. (1991). New developments in pairwise multiple comparisons: Some powerful and practicable procedures. *Psychological Bulletin, 110,* 577–586.

Semel, E., Wiig, E. H., & Secord, W. (1987). *Clinical evaluation of language fundamentals—revised.* San Antonio, TX: Psychological Corporation.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (1992). Identifying the cause of reading disability. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 275–305). Hillsdale, NJ: Erlbaum.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 531–545.

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

Shucard, D. W., Cummins, K. R., Gay, E., Lairsmith, J., & Welanko, P. (1985). Electrophysiological studies of reading disabled children: In search of subtypes. In D. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 87–106). Parkton, MD: York Press.

Siegel, L. S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42,* 201–215.

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22,* 469–479.

Spache, G. D. (1981). *Diagnostic Reading Scales.* Monterey, CA: CTB/McGraw Hill.

Stanley, G., Smith, G. A., & Howell, E. A. (1983). Eye movements and sequential tracking in dyslexic and control children. *British Journal of Psychology, 74,* 181–187.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stanovich, K. E. (1988). Explaining the differences 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., 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.

Tunmer, W. E. (1989). The role of language-related factors in reading disability. In O. Shankweiler & I. Y. Liberman (Eds.), Phonology and Reading Disability: Solving the reading puzzle. Ann Arbor, MI: University of Michigan Press.

Tunmer, W. E., & Hoover, W. A. (1993). Components of variance models of language-related factors in reading disability: A conceptual overview. In R. M. Joshi & C. L. Leong (Eds.), *Reading disabilities: Diagnosis and component processes* (pp. 135–173). Dordrecht, The Netherlands: Kluwer Academic.

Tunmer, W. E., Nesdale, A. R., & Wright, A. D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology, 5,* 25–34.

Vellutino, F. R. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press.

Vellutino, F. R. (1987, March). Dyslexia. *Scientific American, 256,* 34–41.

Vellutino, F. R. (1991). Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. *Journal of Educational Psychology, 83,* 437–443.

Vellutino, F. R., & Scanlon, D. M. (1982). Verbal processing in poor and normal readers. In C. J. Brainerd & M. Pressley (Eds.), *Verbal processes in children* (pp. 189–264). New York: Springer-Verlag.

Vellutino, F. R., & Scanlon, D. M. (1987a). Linguistic coding and reading ability. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (Vol. 2, pp. 1–69). New York: Cambridge University Press.

Vellutino, F. R., & Scanlon, D. M. (1987b). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33,* 321–363.

Vellutino, F. R., Scanlon, D. M., Small, S. G., & Tanzman, M. S. (1991). The linguistic basis of reading ability: Converting written to oral language. *Text, 11,* 99–133.

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. *Journal of Experimental Child Psychology, 59,* 76–123.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1988).

Lexical memory in poor and normal readers: Developmental differences in the use of category cues. *Canadian Journal of Psychology, 42,* 216–241.
Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 279–324). Baltimore: Paul H. Brookes.
Vogel, S. A. (1974). Syntactic abilities in normal and dyslexic children. *Journal of Learning Disabilities, 7,* 103–109.
Wasik, B. A., & Slavin, R. R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28,* 179–200.
Wechsler, D. (1974). *Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation.
Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised.* New York: Psychological Corporation.
Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345–422). Washington, DC: American Educational Research Association.
Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72,* 1–15.
Wolf, M. (1984). Naming, reading, and the dyslexias: A longitudinal overview. *Annals of Dyslexia, 34,* 87–115.
Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised.* Circle Pines, MN: American Guidance Services.
Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised.* Allen, TX: DLM Teaching Resources.
Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher, 49,* 20–29.

Appendix

Administration and Scoring Procedures for Phoneme Segmentation, Decentration, Matching Familiar Figures, and Target Search Tests Administered in Kindergarten, and/or First or Second Grades

Phoneme Segmentation

Initial Phoneme Deletion

Learning items. On this task the child heard a word and was instructed to say the word and then to say it without the initial sound. First the examiner modeled the procedure, using the words *cat* and *bat.* The child was then asked to delete the initial sound in *fat.* If the child was successful, the next practice item was administered. If the child was unsuccessful, the examiner provided corrective feedback as follows:

> Listen to the word again—it's *fat.* It has /f/ at the beginning and then /at/, /f/-/at/. So if you say *fat* without the /f/ it's /at/. Now you try it, say *fat* and then say it without the /f/.

If the child was still unsuccessful, the examiner modeled the task one more time and then asked the child to attempt the word again. A maximum of six items was presented using this format. The first three items rhymed with *at,* and the next three items rhymed with *all.* If the child performed the deletion task correctly the first time an item within a rhyming set was presented, the next item presented was from the other rhyme set. If both sets had been presented, the test items were administered.

Each child received a minimum of two learning items (one from each rhyme set) and a maximum of six learning items. The child's score was based on the number of items he or she attempted and how successful he or she was on those items. Specifically, a child who attempted all six items and was unsuccessful on all of them received a score of 0. A child who attempted all six and was successful on two of them received a score of 2, whereas a child who attempted only two and got them both right received a score of 8, which indicated that he or she was highly successful on the task without having received training or corrective feedback and was thereby credited with 6 correct points and 2 bonus points.

Test items. Following completion of the learning items, the test items were administered by simply asking the child to say each of five words (*cup, sit, chin, feet, bus*) without the first sound. No feedback or modeling was provided for the test items. One point was awarded for each correct response, and the child's score was the total number of items correct.

Final Phoneme Deletion

This subtest was administered only if the child provided at least one correct response on the test items for initial phoneme deletion. The administration procedures for the learning and test items were essentially the same as for the initial deletion task, except that the child was asked to say the stimulus word and then to say it again without the last sound. The items for the learning trials were as follows: *neat, kneel,* and *niece* and *place, plate,* and *plane.* The test items were *hide, make, time, seed,* and *nose.* The learning items were not scored for this subtest. One point was awarded for each correct response on the test items.

Phoneme Articulation

This subtest was administered only if the child had provided at least three correct responses on the final deletion subtest (criterion determined in pilot study). The articulation test was included in the battery to provide enough ceiling on the test.

On this test, the child was asked to articulate the different sounds in minimally contrasted word pairs. For example, for the stimulus pair *nap* and *lap,* the child was asked to say the sounds /n/ and /l/. The administration procedures for this test were similar to those for the phoneme deletion tests in that modeling and feedback were provided for the following learning items: *nap–lap, not–lot, need–lead, mess–miss, pen–pin,* and *set–sit.* No feedback was provided for these test items: *tip–dip, cat–can, mad–made, up–us, net–not,* and *feel–kneel.* Performance on the learning items was not scored. For the test items, partial credit was awarded for an item if a child managed to articulate only one of the sounds that differentiated a given word pair.

Total Score

The final score on the phoneme segmentation test was the sum of the scores for the learning and test items on the initial phoneme deletion component and the scores for the test items on the final phoneme deletion and phoneme articulation components. Scores on the learning items for the initial phoneme deletion component were included to reduce floor effects on the test.

Concrete Operations (Decentration Ability)

Three aspects of decentration ability (Piaget, 1952) were assessed: *conservation, seriation,* and *class inclusion.* Each was measured with two different tasks as described below.

Conservation

The conservation tasks evaluated both number and mass. The number conservation task used plastic circles that were originally arrayed in front of the child in two rows of seven each. The child was told that the two rows had the same number of circles. Then, while the child watched, the examiner moved the circles in the row nearest to the child closer together so that the circles almost touched. The examiner then asked whether the two rows had the same number of circles or a different number of circles. After the child responded, the examiner asked, "How do you know?"

Procedures used for the conservation of mass task were similar. The examiner displayed two round balls of clay that were the same size and told the child that the two balls contained the same amount of clay. Then, while the child watched, the examiner rolled one ball into an elongated shape and asked, "Do these have the same amount of clay or are they different?" After the child responded, the examiner asked "How do you know?" For each conservation task, one point was awarded if the child indicated that the quantities had not changed. An additional point was awarded if the child provided a good reason for his or her response (e.g., nothing was added or taken away).

Seriation

The seriation tasks used sticks in one case and circles in the other. There were six sticks ranging in length from 1 in. (2.54 cm) to $3\frac{1}{2}$ in. (8.89 cm) in $\frac{1}{2}$-in. increments. The six circles ranged from $\frac{1}{4}$ in. (0.64 cm) to 3 in. (7.62 cm) in diameter in approximately $\frac{1}{2}$-in. (1.27 cm) increments. Administration of the two tasks occurred as follows: The sticks (or circles) were placed in front of the child in a horizontal row in a predetermined random order. The child was then asked to "put these sticks [circles] in order from smallest to biggest" as follows:

> Start with the smallest one here. Put the one that is just a little bit bigger right next to it and then the one that is a bit bigger next to that one and keep going until you have used up all of the sticks [circles].

The child received two points for correctly ordering all of the objects in the array. If he or she initially misordered the objects but spontaneously corrected the ordering, one point was awarded.

Class Inclusion

There were two class inclusion items. The first used a 2 × 2 array in which the top left corner contained a red apple, the top right corner a red flower, and the bottom left corner a yellow apple. The bottom right corner was empty. The child was asked to decide which of four pictured objects (red apple, yellow apple, red flower, yellow flower) belonged in the empty box of the array. The second item used a 2 × 3 array with the top row containing a square, a circle, and a triangle, all colored blue. The second row contained a red square, a red circle, and an empty box. The choices for completing this array were a red circle, a red triangle, a blue square, or a blue triangle. For each item, after the child had chosen a picture to complete the array, he or she was asked, "Why did you pick that one?" One point was awarded if the child chose the correct picture and an additional point was awarded if he or she was able to articulate a good reason for making the correct choice.

Matching Familiar Figures Test (MFF)

Kagan's (1965) original version of the MFF is a match-to-sample task in which the child is presented with a target picture and six similar pictures, one of which is identical to the target. The child's task is to identify the identical match. If the child's response is incorrect, the examiner informs him or her of this and asks him or her to choose another item. The child is given up to six trials to find the correct match before the next item is administered. His or her performance is evaluated with both speed and accuracy criteria. Speed is quantified as the time taken to make the first response and was measured with a stopwatch. The accuracy measure is quantified as the number of incorrect responses made. Because we were working with kindergarten children, we simplified the task by reducing the number of options for each item from six to four. Further, the most difficult items from the original test were eliminated altogether. This resulted in a total of seven items.

Administration of the MFF was initiated with two practice items to familiarize the child with the task. The target picture and response choices were presented on opposing pages of a loose-leaf notebook that appeared in a horizontal orientation. The task was introduced as follows:

> I am going to show you a picture of something you know and then some pictures that look like it. You will have to point to the picture on this bottom page that is just like the one on this top page. Look at this picture up here. Now look at these four pictures down here and find the one that looks *just* like the top one. Look very carefully at all of the pictures on the bottom to find the one that looks *just* like the one on top. Look very carefully because some of them are tricky.

Assistance was provided on sample items if the child manifested any difficulty in making the correct choice. For the test items, the examiner used a digital stopwatch to assess time taken to make the first response. Praise was given for each correct response. For incorrect responses, the examiner said, "No, that is not the right one. Find the one that is just like this top one."

Performance indices for the MFF were the mean time to first response and the total number of errors. The MFF was intended to measure the child's ability to attend to visual detail and apply checking and self-monitoring strategies.

Target Search Test (TST)

The kindergarten version of the TST was a modification of the symbol subtest of the target search task used by Rudel et al. (1978). On this task, the child was given an array of geometric shapes presented on an $8\frac{1}{2}$ × 11 in. (21.6 × 27.9 cm) sheet of white paper. The target symbol, a diamond that was approximately $\frac{3}{4}$ in. (1.9 cm) high, was presented at the top center of the page. Below this were 10 lines of eight symbols each. The foils included circles, triangles, and rectangles that were randomly intermixed with the target symbol. The child was instructed to find all occurrences of the target symbol and draw a line through each. He or she was provided with a suggested line-by-line search strategy at the outset of the task.

Three forms of the TST were administered to second graders: symbols, numbers, and letter clusters. For the symbol TST, the target was a diamond, and the distractors triangles, rectangles, circles, squares, and so forth. On the numbers version, a three-digit number (e.g., 529) was the target item and the child searched through an array of target and distractor items (14 rows × 10 columns) to find all occurrences of the target item. Distractor items typically had one or two digits in common with the target item. The letter TST was similar in that the target consisted of a three-letter syllable (e.g., LIF) and the distractors were all three-letter pronounceable syllables that had one or two letters in common with the target. On each subtest, the child was asked to find and draw a line through as many of the target items as possible. To maximize the probability of systematic searching, the child was encouraged to use a left-to-right, row-by-row search strategy. Performance indices for this measure were the time taken to complete the task (measured with a stopwatch) and the number of target items located. Each of the tasks is used to assess planning and organizational skills, in that a systematic search of the display tends to yield the highest hit rate in terms of targets found in the array.

Received April 27, 1995
Revision received April 25, 1996
Accepted May 6, 1996 ■

Instructions to Authors

Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (4th ed.). All manuscripts must include an abstract containing a maximum of 960 characters and spaces (which is approximately 120 words) typed on a separate sheet of paper. Typing instructions (all copy must be double-spaced) and instructions on preparing tables, figures, references, metrics, and abstracts appear in the *Publication Manual.* All manuscripts are subject to editing for sexist language. Manuscript length per se is not an issue, although length should be related to the manuscript's "information value." For further details about appropriate manuscript length and content, authors are referred to the Editorials in the March 1991 issue (Vol. 83, No. 1, pp. 5–7), the March 1992 issue (Vol. 84, No. 1, pp. 3–5), and the March 1993 issue (Vol. 85, No. 1, pp. 3–6) of the *Journal.*

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. In addition, it is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 6.24). As this journal is a primary journal that publishes original material only, APA policy prohibits as well publication of any manuscript that has already been published in whole or substantial part elsewhere. Authors have an obligation to consult journal editors concerning prior publication of any data upon which their article depends. In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 6.25). APA expects authors submitting to this journal to adhere to these standards. Specifically, authors of manuscripts submitted to APA journals are expected to have available their data throughout the editorial review process and for at least 5 years after the date of publication.

Authors will be required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment. A copy of the APA Ethical Principles may be obtained by writing the APA Ethics Office, 750 First Street, NE, Washington, DC 20002-4242.

Because the *Journal* has a masked review policy, authors submitting manuscripts are requested to include with each copy of the manuscript a cover sheet that shows the title of the manuscript, the authors' names and institutional affiliations, the date the manuscript is submitted, and footnotes identifying the authors or their affiliations. The first page of the manuscript should omit the authors' names and affiliations but should include the title of the manuscript and the date it is submitted. Every effort should be made by the authors to see that the manuscript itself contains no clues to their identities.

Authors should submit five copies of their manuscripts. All copies should be clear, readable, and on paper of good quality. A dot matrix or unusual typeface is acceptable only if it is clear and legible. Manuscripts not meeting readability and APA *Publication Manual* specifications will be returned for repair before being reviewed. In addition to addresses and phone numbers, authors should supply electronic mail addresses and fax numbers, if available, for potential use by the editorial office and later by the production office. Authors should keep a copy of the manuscript to guard against loss. Mail manuscripts to the Incoming Editor, Michael Pressley, Department of Educational Psychology and Statistics, University at Albany, State University of New York, Albany, New York 12222. Electronic correspondence to the Incoming Editor can be sent via Internet to jedpsych@cnsibm.albany.edu