

Preventing Reading Failure in Young Children with Phonological Processing

Disabilities: Group and Individual Responses to Instruction

Joseph K. Torgesen
Richard K. Wagner
Carol A. Rashotte
Florida State University

Elaine Rose
Rose Speech and Academic Center

Patricia Lindamood
Lindamood-Bell Learning Processes

Tim Conway
University of Florida and the Morris Center

Cyndi Garvan
University of Florida

Journal of Educational Psychology, In press

The research reported in this manuscript was supported by grant number HD30988 from the national Institute of Child Health and Human Development, and by grants from the national Center for Learning disabilities, and the Donald D. Hammill Foundation. Questions regarding the manuscript should be sent to the Joseph Torgesen, Psychology Department, Florida State University, Tallahassee, FL 32306-1270, e-mail: torgesen@psy.fsu.eduAbstract

Abstract

The relative effectiveness of three instructional approaches for the prevention of reading disabilities in young children with weak phonological skills was examined. Two programs varying in the intensity of instruction in phonemic decoding were contrasted with each other and with a third approach that supported the children's regular classroom reading program. The children were provided with 88 hours of one-to-one instruction beginning the second semester of kindergarten and extending through second grade. The most phonemically explicit condition produced the strongest growth in word-level reading skills, but there were no differences between groups in reading comprehension. Word-level skills of children in the strongest group were in the middle of the average range. Growth curve analyses showed that beginning phonological skills, home background, and ratings of classroom behavior all predicted unique variance in growth of word level skills.

This study was designed to contribute to our understanding of the instructional conditions that need to be in place to prevent reading disabilities in young children. Both the specific design of the study and the questions it addressed were derived from previous research and theory in two areas. The broadest context of the study is the new understanding of reading and reading disabilities we have acquired from research over the past 20 years (Adams, 1990, Metsala & Ehri, 1998), and the more focused context is previous research on instructional methods that accelerate reading development in young children who are either experiencing, or are at-risk for reading failure (Foorman, Fletcher, Francis, Schatschneider, &

Mehta, 1998; Shanahan & Barr, 1995; Wasik & Slavin, 1993; Vellutino, Scanlon, Sipay, Small, Pratt, Chen, & Denckla, 1996).

Within the broader context, perhaps the most important single conclusion about reading disabilities is that they are most commonly caused by weaknesses in the ability to process the phonological features of language (Liberman, Shankweiler, & Liberman, 1989). In particular, individual differences in phonological awareness and rapid automatic naming ability have been shown to exercise unique causal influences on the rate at which children acquire important early reading skills (Wagner, et al., 1997). These two cognitive/linguistic abilities have also been demonstrated to be the most salient disabilities of older children with reading disabilities (Fletcher, et al., 1994; Wolf, 1997).

Discoveries about the core cognitive/linguistic problems of children who experience special difficulties learning to read are important to research on the prevention of reading disabilities for two reasons. First, they provide a means to accurately identify children at-risk for reading disabilities before reading instruction begins (Torgesen & Burgess, 1998). This should allow preventive work to begin earlier in school and focus accurately on children who are most in need of preventive intervention. Second, use of selection criteria involving phonological skills allows identification of a more theoretically coherent sample for study than is frequently the case. Most previous research has focused on children identified for intervention by teacher nomination or socio-economic status (Wasik & Slavin, 1995). While children identified in this manner clearly constitute an "at-risk" group (Bowey, 1995), they are not as cognitively or linguistically coherent as groups identified by specific deficiencies on linguistic measures. The goal of the present study was to examine the effectiveness of several instructional procedures for a specific subset of children who are at-risk for reading difficulties; those who enter school delayed in the development of phonological skill.

The most important impact of weaknesses in the ability to process phonological information is to make it very difficult for children to understand and apply the alphabetic principal in deciphering unfamiliar words in print (Siegel, 1989; Torgesen, in press). These early difficulties in acquiring phonemic decoding skills lead directly to delays in the development of orthographic reading skills which are one important basis of fluent reading (Ehri, 1998; Share & Stanovich, 1995). Children with reading disabilities show word-level reading problems from the beginning of reading instruction, and in the normal course of development, they almost never acquire average level skills in this area (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel, 1988).

Evidence about the central role of word-level reading problems in children with phonetically based reading disabilities suggests that, to be successful with this population, interventions must contain powerful instruction and effective practice at this level. Consistent with this implication, two recent reviews of early intervention research using broadly defined samples of children concluded that the most successful programs to date have included systematic instruction to help children learn to decipher words in print (Pikulski, 1994; Wasik & Slavin, 1993). One study in particular (Iverson & Tunmer, 1993) provided specific evidence that the effectiveness of an early intervention program for at-risk children could be enhanced substantially by the addition of explicit instruction in phonemic decoding skills. However, both reviews also concluded that the most successful interventions were those that were derived from the most inclusive models of reading and which contained an appropriate balance of word and text level instruction, including instruction specifically focused on reading comprehension. Thus, previous research with children identified by teacher nomination or socio-economic status suggests that, to be maximally effective, early intervention programs need to contain a carefully orchestrated mix of instruction to help children construct the meaning of text as well as to read words accurately and fluently.

A potential tension between the needs for instruction designed to build word reading skills and instruction focused on construction of meaning is brought into sharp focus by two recent, carefully

controlled prevention studies. Brown and Felton (1990) contrasted two instructional approaches with a sample of children selected because of weaknesses in phonological development, and Foorman and her colleagues (Foorman, et al., 1998) contrasted three instructional approaches with first grade Title 1 children. In both cases, the instructional condition that contained the most explicit instruction in phonemic decoding skills produced the strongest growth in word-level reading skills, and the Title 1 study also showed parallel differences in reading comprehension across groups. Of particular interest in the present context is the fact that both studies also showed that explicit and systematic instruction in phonemic decoding skills was particularly beneficial for those children who were most impaired in phonological awareness at the beginning of the study. This suggests that, as children's phonological weaknesses become more severe, their need for explicit instruction in word-level skills increases. If the time available for preventive instruction is limited, as it almost always will be, what is the most appropriate balance of instruction for children with severe phonological weaknesses?

In the present study, the most important instructional contrast involved the degree of explicitness of instruction in phonological awareness and phonemic reading skills as well as the extent of decontextualized, focused practice on these skills. Both instructional approaches included in the study were based on the idea that children with phonological processing weaknesses must receive direct instruction in phonemic decoding strategies. However, one approach attempted to create maximum possible strength in phonemic decoding (within the constraints of the amount of instructional time available), while the other approach emphasized the active coordination of less well developed phonemic reading skills with clues from context as a means of accurately reading words in text and constructing meaning. The latter approach was more fully "balanced" in its mix of word and text level instruction.

These two experimental interventions were contrasted with a third intervention that was designed to be more closely coordinated with each child's classroom reading instruction than the other two interventions. This study employed one-to-one tutorial intervention, which is widely regarded as being the most powerful form of instruction for at-risk children (Learning First Alliance, 1998). Furthermore, it makes sense that tutorial instruction that is closely coordinated with the reading instruction the child receives in the regular classroom should have some advantages over instruction that is not coordinated in the same way (Allington & McGill-Franzen, 1989). However, it is also likely that the effectiveness of tutorial instruction will depend upon the nature of the classroom instruction with which it is coordinated. If the classroom instruction contains sufficient elements that are responsive to the needs of children with phonological processing weaknesses, and the tutorial instruction is able to reinforce and explicate this instruction for individual children, outcomes are likely to be better than when the classroom instruction is not specifically responsive to the needs of at-risk children.

One final aspect of the present study derives from the manner in which previous early intervention programs have been evaluated. Typically, these programs have been compared to alternative methods, and found to produce substantially greater gains in reading (Hiebert, Colt, Catto & Gury, 1992; Pinnell, Lyons, Deford, Bryk, & Seltzer, 1994). However, because standardized measures of reading skill are frequently not reported, we often do not know whether student reading skills were in the average range following instruction, or whether they were better than the control group, but still seriously behind in reading. Even when standardized scores are in the average range, individual variability in scores is not explicitly reported, so that it is not possible to determine the effectiveness of the intervention for the most impaired children. Since previous studies have typically selected their samples from the 15-30% most at-risk children, it is clearly possible that overall effects are largely the result of improvements in reading skill among the least learning impaired children. In the present study, we used standardized measures of reading skill to assess overall level of reading outcomes, and we also separately examined performance of our weakest children. The study also employed growth curve methodology to provide information about characteristics of the children who were least responsive to instruction.

Method

Participants

Children were selected to participate in the study using a two tiered screening process. During September, all kindergarten children in 13 elementary schools (N=1436) were given the Letter Naming task which required them to name as many of the 26 uppercase letters as they could. The sample screened was 51% male, 71.9% White, 26% African American, .6% Hispanic, and 1.5 % Asian.

Children scoring in the bottom 30% on the Letter Naming task were given three additional tasks: Phoneme Elision (a measure of phonological awareness), Serial Naming of Numbers (a measure of rapid automatic naming ability), and the Vocabulary subtest of the *Stanford-Binet Intelligence Test, Fourth Edition* (Thorndike, Hagen, & Sattler, 1986). These, and all other tests administered in this study are completely described in an appendix that is available from the first author. Eighteen students were dropped from the second screening because they were repeating kindergarten, were already assigned to receive extensive special education support, spoke English as a second language, or moved.

180 children who obtained the lowest combined scores on the Letter Naming and Phoneme Elision task, and who had an estimated Verbal Intelligence score above 75, were selected for the study. Because so few of the 413 children who received the second screening tests were able to identify number names correctly, there was not an exclusionary criterion based on the Serial Naming task. Children were randomly assigned within school to one of 4 conditions: (1) a no-treatment condition; (2) a regular classroom support (RCS) condition; (3) the embedded phonics (EP) condition; and, (4) the phonological awareness plus synthetic phonics (PASP) (*aka Lindamood; added by CRC*) condition. Each of the 13 schools in which the interventions were delivered had roughly equal numbers of children in each instructional condition.

Descriptive characteristics of children on the screening measures for each of the four groups is given in Table 1. The groups were not significantly different from one another on any of the

TABLE 1: *Descriptive Characteristics of Children in the Four Experimental Conditions*

Variable	No Treatment	RCS	PASP <i>(Lindamood; added by CRC)</i>	EP
Age (months)	66.0 (3.6)	64.9 (3.3)	65.8 (3.8)	65.2 (3.2)
Verbal IQ (est.)	91.8 (9.2)	90.7 (6.8)	92.4 (9.0)	92.7 (10.7)
Letter Name Knowledge	2.5 (1.7)	2.4 (1.7)	2.0 (1.7)	2.4 (1.7)
Phoneme Elision	1.4 (1.5)	1.5 (1.5)	1.3 (1.5)	1.4 (1.6)
Sex Ratio	23M / 22F	26M / 19F	24M / 21F	23M / 22F
Racial Balance	24AF / 21W / 0H	23AF / 21W / 1H	22AF / 22W / 1H	25AF / 20W / 0H
<i>AF=African American</i>		<i>W=White</i>		<i>H=Hispanic</i>

screening variables, although the minority representation in this sample was higher than in the population from which the sample was selected.

Pretests

An extensive battery of pretests was given to the 135 children in the three treatment groups. The battery covered a broad range of cognitive abilities and pre-reading skills in order to provide the basis for estimating child characteristics that might influence response to the educational interventions in the study. Children in the no treatment group were not given the pretest battery, although their reading and phonological development was followed with smaller batteries as the study progressed. The pretests were administered during the months of November and December of the children's first semester in Kindergarten.

The pre-test battery included measures of three types of phonological processing including phonological awareness (Sound Matching, Sound Categorization, and Blending Phonemes), verbal short-term memory (Memory for Digits and Recalling Sentences), rapid automatic naming ability (Rapid Color Naming, Rapid Object Naming, Rapid Letter Naming). We also measured expressive and receptive language using four subtests (Oral Directions, Word Structure, Sentence Structure, and Recalling Sentences) from the *Clinical Evaluation of Language Fundamentals-Revised* (Semel, et al., 1987), general verbal ability (Listening Comprehension, Boston Naming Test), nonverbal ability (Pattern Analysis, Copying), and visual processing skill (Visual Matching, Visual Memory). Finally, we included a variety of measures of reading and reading related knowledge and skill (Word Identification and Word Attack subtests from the *Woodcock Reading Mastery Test-Revised* (Woodcock, 1987), a list of real words and list of nonwords that increased more gradually in difficulty than the standardized tests, and a measure of letter sound knowledge.

We also sent home a general questionnaire about the home literacy environment. Each year we also asked the children's classroom teachers to complete the *The Multigrade Inventory for Teachers* (Agronin, Holahan, Shaywitz, & Shaywitz, 1992) as a measure of academically related behaviors in the classroom. Children in the three treatment conditions were also given the Block Design, Picture Completion, Vocabulary, and Information subtests from the *Wechsler Intelligence Scale for Children-Revised* (Wechsler, 1974) midway through second grade.

Outcome measures

The children's growth in word-level reading skills and phonological awareness was monitored in five assessments following the pre-test. In addition, a more extensive battery of outcome measures was given at the end of second grade, when the interventions were concluded. All tests were administered by research assistants who were blind to the instructional condition in which each child participated. Growth in phonological awareness and word-level reading skills was monitored at the end of kindergarten, mid and end of first grade, and mid and end of second grade with the following tests that had been given at pretest or screening: Phoneme Elision, Phoneme Blending, Word Identification, Word Attack, Real Word List, and Nonword List.

The final assessment at the end of second grade included a broader range of reading tests and measures of spelling and math ability. Additional measures of reading included the Sight Word Efficiency and Phonemic Decoding Efficiency subtests from the *Test of Word Reading Efficiency* (Torgesen, Wagner, & Rashotte, 1999), Passage Comprehension from the *Woodcock Reading Mastery Test-Revised* (Woodcock, 1987) and the *Gray Oral Reading Test-III* (Wiederholt & Bryant, 1992) which provided an additional comprehension score. Measures of spelling skill included the Spelling subtest

from the *Wide Range Achievement Test-Revised* (Jastak & Jastak, 1978), and a Developmental Spelling Analysis (Tangel & Blachman, 1992) which measured accuracy of phonemic representations in spelling. Achievement in mathematics was assessed with the Calculation subtest of the *Woodcock-Johnson Psycho-Educational Battery* (Woodcock & Johnson, 1989).

Interventions and procedures

Children in the treatment conditions were provided with four, 20 minute sessions of one-to-one instruction per week for two and a half years beginning in the second semester of kindergarten. Two of the 20 minute sessions were led by certified teachers, and two were led by aides who followed the teacher's written instructions to practice what the children had learned in the previous day's instructional session. All instruction took place in quiet rooms within schools or in mobile research trailers located on school grounds. Instruction was individually paced according to the child's progress in mastering the skills taught in each condition. Following is a description of each of the instructional conditions:

The Phonological Awareness plus Synthetic Phonics (PASP) program. This group received the Auditory Discrimination in Depth Program as developed and outlined by Patricia and Charles Lindamood (1984). (Underline added by CRC) The program provided explicit instruction in phonemic awareness by leading children to discover and label the articulatory gestures associated with each phoneme. This discovery work was followed with activities to build skills in tracking the phonemes in words using mouth/form pictures, colored blocks, and letters to represent the phonemes in words. Although children in this condition spent most of their time building phonemic awareness and phonemic decoding skills, they also began reading text as soon as they showed reasonably mastery of an initial group of 10 consonants and three vowels. We used the *Poppin Readers* (Smith, 1992) and the *Early Literacy Series* (Hannah, 1993) which have been specially produced to provide phonemically decodable text as children progress through the *Auditory Discrimination in Depth* program.

As soon as children began work on multisyllable words, they did their oral reading in trade books or books they brought from class. Also at this point, they began receiving direct, fluency building practice in recognizing words from a list of words that occurred with high frequency in printed English (Fry, Kress, & Fountoukidis, 1993). During oral reading activities, the children were encouraged to rely on their skills in phonological decoding, but were also encouraged to ask whether the pronunciation of an unfamiliar word "makes sense" in the context of the story. Additionally, the teachers discussed the meaning of passages with the children in order to help focus their attention on the importance of comprehension as the end goal of the reading process.

The Embedded Phonics (EP) program. In its early phases, the instructional periods in this program consisted of four main activities: a) learning to recognize small groups of whole words by using word level drill and word games, b) instruction in letter-sound correspondences in the context of the sight words being learned, c) writing the words in sentences, and d) reading the sentences that were written. Stimulation of phonological awareness was done during writing activities in which the children were asked to identify the beginning and subsequent sounds in words before writing them. Some letter sounds (short vowels, r-controlled vowels) were taught through memorization of picture-word cards, but most grapheme-phoneme correspondences were taught in the context of word reading and writing activities.

As soon as children had a small vocabulary of high frequency words, they began learning words from a basal series that contained short stories made from the words they were learning (HBJ Bookmark Series). Although the stories were read orally after the words in them were learned, the primary emphasis during the entire "basal" phase was on acquiring word level reading skills (sight vocabulary and phonemic decoding skills). Even when reading sentences, however, the children were consistently and explicitly shown how they could combine what they knew about the "sounds" in words with what they knew about the meaning of the sentence in order to help them identify individual words.

After students finished the first grade reader (*Together We Go*) in the basal series, the emphasis of the program shifted from learning to read by writing to learning to read by reading. Student-teacher discussion focused on constructing the meaning of the stories that were read. Although less time was spent on individual word drills, the children continued to receive direct instruction in new "sight words", but rather than being taken from the basal series, their words came from the same list of high-frequency words that was used in the PASP (*essentially the program used by CRC*) program. Writing activities during this phase of instruction continued to include generation of sentences containing specific words, but also included more free writing and writing summaries of stories that were read.

The Regular Classroom Support Condition (RCS). The children assigned to this condition received individual tutoring in the activities and skills taught in their regular classroom reading programs. The activities provided to children in this group necessarily varied somewhat across the different schools in the study because of the slightly different orientations toward reading instruction of the regular classroom teachers involved in the study. The activities varied from "phonics" oriented activities such as explicit instruction in letter-sound correspondences, to intermediate activities involving sight word drill, to activities focusing on meaningful communication such as writing in journals or discussing stories that the teacher read. A more specific time by activity breakdown will be provided when the activities of all three instructional conditions are compared.

Selection and training of tutors and aides

The nine project tutors had varying degrees of experience teaching young children. At the time they were interviewed and selected to participate in the project, each of the tutors in the project had been rated by the project director on a five point scale in three areas related to experience and expertise in teaching reading. The tutors were matched in pairs on their total ratings, and one member of each pair was randomly assigned to either the EP or PASP (*essentially the program used by CRC*) conditions. The ninth tutor, whose total rating score was very close to the overall average, was assigned to the PASP condition.

Tutors initially received 18 hours of training in the instructional program to which they were assigned. Training for tutors in the PASP (*essentially the program used by CRC*) program was provided by Patricia Lindamood, and tutors in the EP program were trained by Elaine Rose, who has used methods very similar to those in the EP program in University based and private clinical practice for approximately 15 years

Each of the tutors was assigned to instruct children in the RCS condition as well as the EP or PASP (*essentially the program used by CRC*) program. However, since the RCS condition involved following the instructional suggestions of each child's regular classroom teacher, no special training was provided relevant to this condition. The tutors had an initial meeting with the child's regular classroom teacher in which the classroom reading program was discussed, and ways in which the project tutor could support that program were identified. Following this initial meeting, the project tutors met with the regular classroom teachers on a weekly or bi-weekly basis (depending upon teacher preference) to discuss specific instructional activities for children in the RCS condition. All tutors were carefully instructed not to use specific techniques from either the EP or PASP (*essentially the program used by CRC*) program in teaching children in the RCS group, unless those procedures were part of the child's regular classroom reading program.

Once instruction began, our tutors met bi-weekly with their project consultant and the project director and coordinator for a three hour inservice training period (the PASP (*essentially the program used by CRC*) and EP groups met separately). During the week prior to these meetings, each tutor was videotaped teaching one child, and these videotapes were used by the consultants to identify areas in which the tutors needed to modify or improve their procedures.

Instructional aides for the project were either undergraduate students in psychology working for course credit, individuals already working in the schools as aides and assigned part-time to the project, or individuals paid directly by the project. Although each half-time tutor was responsible for providing instruction to approximately 15 children, the number of children seen by each aide ranged from 2 to 15. As with the tutors, the aides were usually assigned to work both with children in the RCS condition, and in either the EP or PASP (*essentially the program used by CRC*) condition. Aides were randomly assigned to either the PASP or EP condition, and the number of aides from each category (students, school-based, paid) assigned to each condition was roughly equivalent.

All aides received two hours of preservice training, along with a manual that described the instructional philosophy and goals of their program and descriptions of specific instructional and practice activities they would be providing. Once instruction began, the aides received a two hour inservice once each month, which was conducted by the project director. The aides were guided in their daily instructional activities by directions left by the tutor in the Instructional Log that each of them kept.

Description of classroom reading instruction

As a rule, we tried to schedule children for our instructional interventions at a time in their school day that did not interfere with their regular classroom reading instruction. We were successful with this scheduling arrangement about 75% of the time, and whether or not our instruction took place during a child's primary classroom reading instruction varied from semester to semester during the two and a half years of the project. Since all of the children in this study received the bulk of their reading instruction in the regular classroom, we assessed the nature of this instruction through the use of teacher surveys. There were no significant differences across groups in the reading approach taken by classroom teachers, and the self-ratings were consistent with our observations that regular classroom reading instruction was primarily literature based and guided by a whole language philosophy, with phonics being taught on an as-needed basis rather than systematically.

Instructional time and time by activity analyses

Our plan was to provide four, 20 minute instructional sessions per week for five school semesters beginning in the second semester of the kindergarten year. This would have allowed us to provide 60 hours of tutor-led, and 60 hours of aid-led instruction for each child. However, because of holidays, field trips, assemblies, student and instructor absences, actual hours of teacher-led instruction in the RCS, PASP (*essentially the program used by CRC*), and EP groups was 46.8, 48.1, and 46.7, respectively. Hours for instruction led by aides was 41.5, 41.9, and 40.4, respectively. The number of hours of instruction by teachers was higher than for aides because we attempted to provide a substitute for tutors when they were ill, but we could not provide substitutes for aides. There were no significant differences across instructional conditions in the number of either type of session, and the total average instructional time across conditions was about 88 hours (47 hours from tutors and 41 hours from aides).

Time spent on various types of instructional activity was monitored in two different ways across conditions. Approximately one of every four tutor led sessions for the PASP (*essentially the program used by CRC*) and EP programs was video-taped for supervision purposes. In the RCS condition, tutors described the instructional activities that occurred within each session and classified them into one of four categories. They did not indicate the amount of time spent on each activity, but simply whether or not the activity was part of the session. Thus, for the PASP (*essentially the program used by CRC*) and EP programs, it was possible to estimate the percentage of instructional time spent on each activity, and for the RCS condition, it was possible to determine the frequency with which different types of instructional activities occurred across sessions.

Although activities occurred in the PASP (*essentially the program used by CRC*) condition that did not occur in the EP condition, and vice versa, it was possible to identify three general categories of instructional activity that were roughly comparable across conditions. The first category involved activities designed to stimulate phonological awareness, teach letter-sound correspondences, and practice reading and writing individual words in which the emphasis was on phonemic decoding. Over the course of the study, tutors in the PASP (*essentially the program used by CRC*) condition spent 74% of their time on this type of activity, versus 26% for the EP condition. The second category involved direct instruction and practice in acquiring sight words. Time spent on this activity in the PASP (*essentially the program used by CRC*) condition was 6%, while for the EP condition, it was 17%. The last category involved reading or writing connected text, which included sentence to book level activity. Twenty percent of the time in the PASP condition was spent on this activity, compared to 57% in the EP condition. From these estimates, it is clear that children in the PASP condition spent much more time in building phonological awareness and phonemic decoding skills, while children in the EP condition spent more time acquiring sight words and engaging in meaningful experiences with print.

The tutors in the RCS condition were asked to classify their instructional activities in each session into one of four categories. The categories were: (1) sight word learning; (2) phonological awareness and phonemic decoding; (3) spelling; and, (4) meaning-emphasis activities with connected text. Over the course of the study, the total frequency of occurrence of these different types of activity within sessions was 24% for sight word activities, 24% for phonics activities, 9% for spelling activities, and 43% for meaning-emphasis activities. Thus, although the specific activities focused on different content, the overall distribution of instructional focus for the RCS program was more similar to the EP program than to the PASP (*essentially the program used by CRC*) program.

Results

Preliminary Analyses

Over the two and a half years of the study, we experienced a 23% drop out rate, so that the sample at the end of second grade contained 138 children. The number of children remaining in each condition was 32, 37, 33, and 36 for the control group, RCS, PASP (*essentially the program used by CRC*), and EP groups, respectively. The only children who were lost from the study were those who moved to distant communities, since we followed children as they changed schools locally. When we began the study, the children were attending 13 elementary schools, and by the end of the study, they were scattered among 23 different schools.

Twenty six percent of the sample had been retained in either kindergarten or first grade, and there was a significant difference in retention rate across conditions, $\chi^2(3) = 8.7, p < .05$, with the percentages for the no treatment group, RCS, EP, and PASP (*essentially the program used by CRC*) conditions being 41%, 30%, 25%, and 9%, respectively. The percentage of children who were referred for special services in the schools, either special education or Title 1 reading programs, also differed (NTC = 22%, RCS = 24%, EP = 42%, and PASP = 18%). Although the difference was not statistically significant in the overall comparison, $\chi^2(3) = 5.8, p > .05$, if just the two experimental conditions (PASP vs. EP) are compared, the difference in referral rate for special services is very reliable, $\chi^2 = 13.7, p < .01$.

The difference in retention rate across conditions presents two kinds of problems for determining the instructional effects of our interventions. If first and second grade children are combined together in the analyses, then the different conditions would contain disproportionate numbers of children who had not received second grade level classroom instruction. This would constitute an obviously confounding difference in the "background" instructional experiences of children across conditions. However, it could be argued that second grade level instruction would be too difficult (and hence unprofitable) for

the children who were retained, and comparisons involving all the children provides the most realistic assessment of the "practical" effects of each instructional condition.

The obvious alternative to using all children in the analyses would be to equate conditions on background instructional experiences by using only children who had completed second grade at the end of the study. However, this type of analysis would seriously underestimate the differential effects of the instruction provided to the PASP (*essentially the program used by CRC*) group, because a higher percentage of children with the most serious learning problems (assuming that was the cause of their retention) would be eliminated from the other groups. For example, because of the high proportion of retentions in the no treatment control group, there were only 19 children in this group in second grade at the end of the study. What would it mean to compare the scores of this select group of children (very likely the most intact learners in the group) to a much more complete group of children from the PASP condition?

Comparisons of groups on outcome measures

As a compromise to these two extremes, we will first present summary data on a small group of the most important outcome measures for all children in each group so that the relative effects of the interventions can be assessed with all children. Following that, we will present a more detailed analysis of differences in outcomes for our two experimental conditions using groups of children that were equated for background instructional experiences.

Table 2 presents outcomes at three measurement points for word level reading (end of kindergarten, first, and second grade for most children), at two measurement points for reading comprehension (end of first and second grades), and at one point (end of second grade) for spelling and math skills. The outcomes with multiple measurement points were analyzed using the

TABLE 2: Reading Outcomes for All Children at the End of Kindergarten, First, and Second Grades

MEASURE	GRADE	GROUP			
		<u>NTC</u>	<u>RCS</u>	<u>PASP</u> (<u>Lindamood</u>)	<u>EP</u>
Word Attack	K	.13 (.50)	.14 (.53)	.76 (1.7)	.28 (1.3)
	1	2.8 (4.0)	4.9 (6.6)	9.8 (8.1)	4.8 (5.2)
	2	10.4 (8.7)	12.2 (10.6)	21.3 (11.1)	12.0 (8.8)
Nonword List ¹	K	.48 (1.3)	.57 (1.4)	1.8 (3.1)	.61 (1.9)
	1	5.9 (9.2)	8.3 (11.2)	19.2 (14.4)	7.5 (7.3)
	2	23.2 (22.4)	23.1 (22.9)	43.2 (20.0)	26.3 (20.8)
Word Identification	K	1.1 (2.5)	1.6 (2.0)	1.2 (1.6)	2.7 (3.1)
	1	16.6 (14.4)	21.4 (14.7)	25.6 (15.8)	20.3 (13.4)
	2	37.2 (17.2)	39.6 (16.0)	47.9 (16.8)	40.9 (14.4)

Real Word List ²	K	4.5 (5.4)	5.6 (4.5)	5.1 (5.3)	7.5 (7.0)
	1	40.8 (36.8)	50.5 (35.8)	65.8 (34.9)	48.6 (33.2)
	2	103.9 (51.5)	109.7 (51.1)	137.9 (43.0)	116.5 (43.1)
WRMT Passage Comp.	1	6.8 (7.1)	9.4 (6.8)	10.7 (7.7)	8.4 (7.1)
	2	19.3 (9.1)	18.9 (9.6)	23.2 (9.6)	19.5 (9.0)
GORT - III Comp.	1	2.2 (3.5)	3.1 (3.3)	3.8 (3.4)	3.7 (4.8)
	2	8.8 (6.6)	8.1 (5.4)	10.9 (7.2)	11.0 (7.6)
WRAT Spelling		30.5 (5.5)	30.9 (5.9)	33.5 (5.7)	31.5 (4.4)
Developmental Spelling		22.5 (4.2)	22.5 (5.2)	25.5 (3.7)	22.5 (5.6)
Calculation		7.8 (2.9)	7.3 (2.9)	8.2 (2.6)	7.7 (2.8)

Note: Preliminary reports on the first and second grade outcomes for some of these measures were presented in earlier discussions of this study that appeared as part of special issues in *Scientific Studies of Reading* (Torgesen, Wagner, & Rashotte, 1997), and *Learning Disabilities: An Interdisciplinary Journal* (Torgesen, Wagner, Rashotte, Alexander, & Conway, 1997).

¹*Non Word List*. A list of 63 nonwords that increased in difficulty more gradually than the Word Attack Subtest.

²*Real Word List*. A list of 104 words that increased in difficulty more gradually than the Word Identification Subtest.

MANOVA procedure from SAS (SAS Institute, 1997) software. Measures of similar constructs were combined together in the analyses, and all three reading outcomes (phonetic decoding, real word identification, reading comprehension) showed significant growth across measurement points, F 's (4, 129) = 69.8, 268.9, and 246.7, respectively. For the measures of phonemic decoding skill (Word Attack and Nonword List), both the effects of group F (6, 264) = 4.7, and the interaction between group and time F (12, 393) = 3.0, were statistically reliable ($p < .001$). Simultaneous individual contrasts showed that the PASP (*essentially the program used by CRC*) group was reliably different ($p < .05$) from the other groups, who did not differ among themselves.

The groups also performed differently on the measures of real word reading (Word Identification, Real word list), F (6, 264) = 2.3, $p < .05$, but the interaction between group and time was not significant. Among the groups, the PASP (*essentially the program used by CRC*) group performed significantly better than the control group and the RCS group, but was not reliably different from the EP group. Performance on the comprehension measures was also reliably different across groups F (6, 264) = 2.4, $p < .05$, however the individual contrasts on these measures did not indicate reliable differences between any of the specific groups. Of the measures that were assessed at the end of the intervention, only the developmental spelling test showed significant differences across groups F (3,133) = 3.1, $p < .05$. For this

measure, the only individual contrast that was reliable was between the PASP and the no treatment control group.

To determine the absolute level of effectiveness of our interventions, we compared the performance of children in our sample against that of children in the standardization samples of the *Woodcock Reading Mastery Test* and the *Gray Oral Reading Test*. Age-based standard scores (mean = 100, S.D. = 15) for the PASP (*essentially the program used by CRC*) group were 99.4, 98.2, and 91.5 for Word Attack, Word Identification, and Reading Comprehension (average of the two comprehension tests), respectively. For the EP group, these scores were 86.7, 92.1, and 88.6, and for the RCS group they were 86.7, 92.1, and 88.3. Scores for the no treatment control children on the same measures were 81.6, 86.3, and 85.8. Thus, children in the intervention condition (PASP) that most consistently obtained the highest reading scores in cross group comparisons also performed at very close to average levels on word level reading skills, and at the low end of the average range in reading comprehension.

In order to provide a comparison of outcomes for the PASP (*essentially the program used by CRC*) and EP conditions unconfounded with differences in background reading instruction, we eliminated the five children from the second grade PASP group who were most closely matched at the end of kindergarten to children from the EP group who had been retained. Since nine of the remaining 36 children (25%) in the EP group had been retained, and only three of the remaining 33 children (9%) from the PASP group had been retained, by eliminating an additional five children from the PASP group, we roughly equated the percentage of "weaker" children eliminated from each remaining sample. We selected the children to be eliminated from the PASP group by first determining that retained and non-retained children in the EP group were most different from one another at the end of kindergarten on measures of word reading, letter name knowledge, and letter sound knowledge. We matched the three retainees from the PASP group to three retainees from the EP group on these measures, and then found five other children from the PASP group who matched as closely as possible other retainees from the EP group.

A comparison of the children in our "matched" sample on the pretest measures is presented in Table 3. Some of the variables in Table 3 were constructed according to the results of a confirmatory factor analysis we performed on these data. This analysis indicated that we could

 TABLE 3: Descriptive Characteristics of Second Grade Childre in the Matched Sample

<u>Variable</u>	<u>GROUP</u>	
	<u>PASP</u> (n=25) (<i>Lindamood</i>)	<u>EP</u> (n=27)
Verbal Ablility	103.3 (12.2)	101.0 (11.5)
Nonverbal Ability	103.6 (10.1)	100.0 (11.5)
Home Background	102.9 (12.1)	100.1 (14.7)
Phonological Awareness	100.2 (8.3)	99.3 (8.4)
Phonological Memory	100.5 (14.4)	99.1 (12.8)
Rapid Automatic Naming	103.0 (13.3)	104.5 (14.5)
WISC-R Estimated IQ	89.7 (12.0)	86.7 (14.0)
Sex Ratio	13M / 11F / 1H	10M / 17F

group the following indicators together to measure: 1) *Verbal Ability* - Boston Naming Test, Sentence memory, Vocabulary from the Stanford Binet, Word Structure, and Listening Comprehension; 2) *Nonverbal Ability* – Pattern Analysis, Copying, Visual Memory; 3) *Home background* – father’s and mother’s education and occupations (scored according to criteria described in Hollingshead (1975). These indicators were standardized on the entire pre-test sample and combined into a unit-weighted composite for each construct. This standard score was then transformed to have a mean of 100 and a standard deviation of 15 for the entire sample at pre-test. In the same manner, the following indicators were also combined to form summary variables: 1) *Phonological Awareness* – Elision, Phoneme Blending, Sound Matching; 2) *Phonological Memory* – Memory for Digits, Sentence Memory, 3) *Rapid Automatic Naming* – Naming Colors, Naming Objects. The WISC-R IQ score was estimated from performance on the four subtests of that measure that were given to the children in second grade. The groups were not reliably different from one another on any of the pretest or demographic variables.

The performance of the matched groups of children in the PASP (*essentially the program used by CRC*) and EP conditions on measures of phonemic awareness and reading are presented in Table 4. Tests of phonemic awareness and untimed measures of phonetic decoding and word identification were administered six times from the pretest to the end of second grade. Measures of fluency of phonetic decoding and word identification were administered at the end of first grade, and at mid and end of second grade. Finally, reading comprehension was measured at the end of first and second grades.

TABLE 4: *Growth in phonemic awareness and reading from pre-test to end of treatment for children in the PASP (essentially the program used by CRC) and EP groups matched for percent of sample removed because of grade retention.*

<u>Outcome Measure</u>				<u>Measurement</u>	<u>Point</u>		
<u>Phonological Awareness</u>		1	2	3	4	5	6
Elision	PASP	1 (2)	6 (2)	10 (4)	12 (4)	14 (4)	17 (5)
	EP	1 (2)	5 (3)	8 (4)	10 (4)	13 (5)	13 (4)
Blending	PASP	3 (2)	7 (4)	13 (4)	16 (3)	18 (4)	19 (5)
	EP	3 (4)	5 (4)	10 (5)	14 (4)	17 (3)	18 (3)
<u>Phonetic Decoding</u>							
Word Attack	PASP	0 (0)	1 (2)	7 (5)	12 (8)	18 (9)	24 (10)
	EP	0 (0)	0 (2)	2 (4)	6 (5)	10 (8)	14 (9)
Nonword List	PASP	0 (0)	2 (3)	12 (9)	23 (14)	34 (16)	50 (17)
	EP	0 (0)	1 (2)	3 (4)	10 (7)	17 (12)	31 (21)

Real Word Reading

Word Ident.	PASP	0 (0)	2 (2)	14 (8)	31 (15)	46 (13)	55 (13)
	EP	0 (0)	3 (3)	12 (9)	25 (12)	37 (12)	46 (11)
Real Word List	PASP	1 (1)	6 (6)	33 (20)	78 (32)	120 (39)	154 (30)
	EP	1 (1)	9 (8)	23 (22)	60 (30)	95 (35)	131 (35)

Fluency Measures¹

Phonemic Decoding Eff.	PASP				12 (8)	18 (12)	25 (13)
	EP				6 (5)	10 (7)	14 (10)

Sight Word Efficiency

PASP	22 (13)	33 (15)	43 (15)
EP	17 (10)	27 (12)	36 (14)

Reading Comprehension

WRMT	PASP		13 (8)	26 (8)
	EP		11 (6)	22 (8)
GORT-III	PASP		5 (3)	13 (7)
	EP		3 (5)	13 (8)

¹Fluency Measures – these measures of fluency of phonetic decoding and sight word reading require children to pronounce as many nonwords or real words as possible in 45 seconds from lists that gradually increase in difficulty. The score is the average between two forms, and reliability is above .90 for both tests.

Similar to the analyses involving children in all four groups, multivariate analyses of variance with repeated measures were used to test for differences between groups on conceptually similar measures, and again these analyses showed significant growth on all measures across measurement points. Differences between groups for phonemic awareness, untimed phonemic decoding, and untimed real word reading were all statistically significant ($p < .05$), with F 's (1,49) = 5.78, 21.6, and 4.83, respectively. Only for phonemic decoding was there a reliable interaction between group and time $F(10,40) = 3.42, p < .05$.

The fluency measures were subjected to univariate repeated measures analyses using the MIXED procedure from SAS (SAS Institute, 1997). Both the effect of group, $F(1,49) = 12.0$, and the interaction between groups and time, $F(2, 98) = 3.64$ were statistically significant for phonemic decoding efficiency. The groups were not significantly different on the sight word efficiency test, nor were they different from one another in the multivariate analyses of the reading comprehension measures.

Because we have given the phonemic decoding efficiency and the sight word efficiency tests to large groups of randomly selected children in the school district in which this study took place, we were able to calculate standard scores on these measures (Mean = 100, S.D. = 15) for children in the PASP (*essentially the program used by CRC*) and EP groups. The average standard scores for phonemic decoding efficiency for the PASP and EP groups were 105.9 and 93.4, respectively, and the scores for sight word efficiency were 99.4 and 94.2.

Individual differences in response to the interventions

We turn now to a consideration of individual differences in response to the interventions. Even though the scores of children in our most effective condition were within the average range on all reading measures, there was still a substantial proportion of the children whose word level reading skills remained relatively unaffected by the intervention. Children performing one standard deviation below average are generally considered to be significantly impaired in reading skill. Table 5 presents the percent of children (including children in both first and second grades) in each condition who performed more than one standard deviation below average for their age on the

TABLE 5: *Percentage of Children in Each Instructional Condition Performing Significantly Below Average and Above Average in Reading Skills at the End of Second Grade.*

	<u>GROUP</u>							
	<u>CONTROL</u>		<u>RCS</u>		<u>PASP</u>		<u>EP</u>	
	85<	>100	85<	>100	85<	>100	85<	>100
Word Attack	53	4	44	17	24	42	47	19
Word Identification	53	25	31	28	21	47	28	22
Passage Comprehension	56	16	56	22	36	36	47	14

Word Attack, Word Identification, and Reading Comprehension subtests (combined score on WRMT-R and GORT-III) as well as the percent who performed above average (standard score above 100). Even in our strongest condition, 24% of this sample of highly at-risk children remained significantly impaired in phonemic reading skills, and 21% remained impaired in real word reading ability.

In the next set of analyses, we attempt to identify the factors most responsible for the variability in reading growth discussed in the preceding paragraph. The technique used involved calculating individual growth curves for measures of word level reading skills (untimed measures of phonetic decoding and real word reading), and then examining the child level characteristics associated with variability in growth parameters. All children remaining in the sample at the end of the intervention were used in this analysis.

Hierarchical linear modeling was used for the growth curve analysis. Five equally spaced measurement points were available (pretest scores could not be used, because they had insufficient variability), and the intercept was fixed at the final measurement point to represent status at outcome. Linear and

quadratic terms were sufficient to model growth in phonetic decoding, as represented by the Word Attack test. Linear, quadratic, and cubic terms were sufficient to model growth in real word identification, as represented by the Word Identification test. For both outcome measures, substantial and reliable individual differences across subjects were obtained for the intercept and linear components of the models, but not for the quadratic or cubic components. Consequently, the intercept and linear components were treated as random effects (i.e., parameters that were free to vary for each subject) and the quadratic and cubic components were treated as fixed effects (i.e., the group average parameter was used for each subject).

Results presented in Table 6 address the issue of what predicts variability in response to training as represented by individual differences in growth curve parameters across subjects. The

TABLE 6: *Individual and Simultaneous Prediction of Growth Curve Parameters for Word Analysis and Word Identification.*

	<u>WORD ATTACK</u>			
	<u>Individual Prediction</u>		<u>Simultaneous Prediction</u>	
	Bivariate Regression	Coefficients	Multiple Regression	Coefficients
	Intercept	Linear	Intercept	Linear
<u>Phonological Variables</u>				
Awareness	.412**	.098**	.241*	.065*
Naming	.346***	.082***	.216**	.045**
Memory	.244**	.058**	.188*	.049**
<u>Cognitive Variables</u>				
Verbal Ability	.300***	.071***	-.117	-.038
Nonverbal Ability	.170*	.040*	-.082	-.021
<u>Noncognitive Variables</u>				
Home Background	.304***	.072***	.241***	.061***
Reading Experience	.089	.021	---	---
Type of Instruction	-.128	-.030	---	---
Behavior Rating	.313***	.074***	.190**	.051**

**Percentage of
Parameter
Variance
Accounted
for:**

36%

37%

**WORD
IDENTIFICATION**

Individual Prediction

Simultaneous Prediction

Bivariate
Regression Coefficients

Multiple
Regression Coefficients

Intercept Linear

Intercept Linear

**Phonological
Variables**

Awareness .469* .098*
 Naming .669*** .140***
 Memory .316* .067*

.044 .001
 .384*** .070**
 .150 .042

**Cognitive
Variables**

Verbal Ability .519*** .109***
 Nonverbal Ability .353* .074*

.004 -.003
 -.094 -.022

**Noncognitive
Variables**

Home Background .497*** .104***
 Reading Experience .233 .049
 Type of Instruction -.134 -.028
 Behavior Rating .681*** .143***

.360** .083**
 --- ---
 --- ---
 .495*** .109***

**Percentage
of Parameter
Variance
Accounted
for:**

45%

45%

first two columns contain regression coefficients for phonological, cognitive, and noncognitive variables considered one at a time (i.e., with each variable as a single predictor in a separate regression model). The tests used to form the composites for phonological awareness, naming, memory, verbal ability, nonverbal ability, and home background are the same as those used to provide the pretest comparison between the PASP and EP groups reported in Table 3. *Reading Experience* was estimated from the *Survey of Home Literacy Environment* by combining parental responses to questions asking about amount of shared reading between parent and child, parental reading habits, and number of magazine subscriptions in the home. *Type of instruction* was meant to describe the balance between phonics/whole language instruction in the child's regular classroom reading instruction during first and second grades. It was the average of teacher self-reports from each child's first and second grade teacher. Finally, *Behavior Rating* was the average score from the child's kindergarten teacher on the categories of activity level, attention, adaptability, social behavior from the *Multigrade Inventory for Teachers*.

Note that because the independent variables were standardized composites, comparisons of the magnitude of regression coefficients across each predictor variable and within each outcome measure are meaningful for each of the growth curve parameters. However, comparisons of the magnitudes of the regression coefficients between growth curve parameters (i.e., intercept versus linear) are not meaningful, nor are comparisons between the two outcome variables because these variables were not standardized. In general, individual differences in response to training were predicted reliably by each of the predictor variables with the exception of reading experience in the home and type of instruction in the classroom.

The second two columns of Table 6 contain regression coefficients for the predictors considered simultaneously (i.e., with the set of variables as predictors in a single regression model). These regression coefficients address the issue of whether the contributions to prediction made by the predictor variables are independent of or redundant with that of the other predictor variables. All predictors with significant regression coefficients when considered individually were included in the simultaneous analysis.

For both measures of word-level reading skill, the most consistently important variables were rapid naming, home background, and classroom behavior ratings. A broader array of phonological variables contributed uniquely to the prediction of growth in word attack skills than was the case for real word identification skills. For both kinds of word reading skill, general verbal ability did not remain a significant predictor of growth once the other predictors were in the equation. We also tested whether the set of unique predictors was reliably different for children who received reading instruction by different methods, and none of the predictor by instructional condition interactions was statistically reliable.

Because the issue of the relationship of general intelligence to the growth of early word reading skills is so important to diagnostic issues in the area of reading disabilities (Fletcher, et. al, 1994), we examined this issue more closely in the growth data. For example, general verbal ability was a reliable predictor of growth in both phonemic decoding and real word reading ability when considered by itself. However, for both types of reading outcome, general verbal ability did not contribute uniquely to prediction of growth once the other variables were in the equation. One obvious possibility is that the presence of the SES variable in the equation is masking the effects of general verbal ability, since these two variables are correlated ($r = .49$). When the SES variable is removed from the predictive equation for phonemic and real word reading ability, the coefficient for verbal ability remains nonsignificant ($p > .05$). For both reading outcome variables, it is not until all the phonological variables are removed from the equation that verbal ability makes a "unique" contribution to the prediction of growth. Since, as in previous research (Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993), all of the phonological variables were significantly correlated with the verbal ability measure (r 's ranged from .21 to .51), it is apparent that the correlation of general verbal ability with growth in word level reading skills is mediated

exclusively by its phonological component. This suggests that the type of verbal ability required in the growth of word level reading skills is primarily phonological in nature.

Since we had only two measurement points for reading comprehension, we were not able to conduct a similar analysis using growth curves to identify the variables most predictive of individual differences in comprehension outcomes. However, when we used the same predictor variables in a standard regression analysis with second grade comprehension scores as the dependent variable, general verbal ability did emerge as a unique predictor, even when the phonological variables remained in the equation. In the case of reading comprehension, variables that uniquely predicted individual differences in outcome were rapid naming, behavior ratings, and either general verbal ability or SES. When these latter two variables were included together in the prediction equation, neither accounted for significant unique variance in the outcome.

General Discussion

This study examined the effectiveness of three approaches to tutorial instruction with a sample of young children selected in kindergarten because they were among the bottom 12% in phonological processing skill. The two experimental approaches varied in their relative emphasis on explicit and intensive instruction in phonemic decoding skills. The third condition, the regular classroom support group, was more closely coordinated with regular classroom instruction than either of the other two.

One of the most important questions of the study concerned the balance between word and text level instruction most appropriate for children with severe phonological weaknesses. Teachers in the PASP (*essentially the program used by CRC*) condition spent 80% of their time on word level instruction and 20% on text level activities, while teachers in the EP condition spent 43% and 57% on word and text level activities, respectively. Analysis of instructional outcomes showed that the PASP group had significantly stronger skills than the EP group in phonological awareness, phonemic decoding, and untimed, context-free word reading. In the analyses involving all the instructional conditions, children in the PASP group were also stronger on word-level reading skills than children in the RCS and the no treatment control group. Although the analysis involving all the groups showed an overall effect of the interventions on reading comprehension, the effect was not strong enough to show reliable differences among specific groups. In the more powerful direct contrast between the EP and PASP groups, there was also no reliable difference in passage comprehension between groups.

In terms of our original question about the appropriate balance between word and text-level instruction for children with serious phonological weaknesses, it is possible to view these results in either of two ways. On the one hand, one could argue that the PASP (*essentially the program used by CRC*) condition was the most effective because it was the only instructional regimen that produced a reliable effect on the word-level reading skills of this group of highly at-risk children. Neither the RCS nor EP instructional conditions produced growth in word-level skills that was reliably different from the no treatment control group. This finding suggests that one-to-one instruction in reading, even if it is consistent with the regular classroom instructional program, may not have a significant impact on the core word-level reading problems of children with serious reading disabilities unless it contains very explicit and intensive instruction in phonemic awareness and phonemic decoding skills. We should not assume that even skillfully administered one-to-one instruction will have a significant impact on word-level skills in children who have serious phonological processing weaknesses if it does not contain sufficient depth of instruction in alphabetic reading skills.

On the other hand, the ability to construct the meaning of written text is the most important outcome from reading instruction, and we found no evidence that children in the three instructional groups were reliably different from one another on this variable. Without evidence for differences in comprehension, it is not possible to assert that any one of the instructional approaches in this study was ultimately more

effective than the other. In order to refine our thinking about these two alternative interpretations of the present results, it is useful to consider plausible explanations for the failure to find comprehension differences among our intervention groups.

We should first point out that these results are very similar to those from the only other prevention study that selected children for intervention who were specifically impaired in phonological processing skills. Brown and Felton (1990) found significant advantages in word-level reading skills for children in their most phonemically explicit condition, but the differences in reading comprehension were not statistically reliable. Among studies that have selected children for intervention based on SES or teacher nomination, effects are consistently larger for word-level skills than for reading comprehension (Shannahan & Barr, 1995; Wasik & Slavin, 1993), although a number of studies have found significant effects on comprehension measures (Foorman, et al., 1998; Hiebert, et al., 1992; Slavin, Madden, Karweit, Livermon, & Dolan, 1990). In each of these latter cases, the most effective interventions have been more broadly focused than our PASP (*essentially the program used by CRC*) condition. That is, they have contained systematic, well designed instruction in both word-level *and* comprehension skills. Perhaps these latter interventions were effective in improving comprehension performance *not* because of their impact on word-level skills, but because they contained more effective instruction than the control conditions in *both* word-level and comprehension skills. In this view, we might consider the PASP condition to be lacking sufficient depth of instruction and practice in applying word level reading skills to constructing meaning. At the same time, we would also recognize that the EP and RCS conditions did not contain sufficient depth of instruction in alphabetic reading skills to make a difference in this area, which may have important implications for future reading growth. Since we plan to follow the development of reading skills in these children for two more years, we may be able to answer questions about this latter possibility.

At first glance, the present findings might be taken to undermine the importance of word-level reading skills as they contribute to good reading comprehension (Gough, 1996). However, we know from previous research that children can be taught processing skills to improve comprehension *independently* of improvements in word-level skill (Mastropierie & Scruggs, 1997). The EP children spent much more time than the PASP (*essentially the program used by CRC*) children reading connected text, answering questions, and discussing meaning with their tutors than did children in the PASP condition (57% vs. 20%). These extra comprehension activities and instruction may have provided the means for these children to construct the meaning of passages as well as children in the PASP condition, in spite of their weaker word-level skills. This possibility is given support when we consider the size of the between-group difference in word level skills (Word Attack = 104 vs. 90, Word Identification = 103 vs. 94) compared to the overall range of these skills within groups (Word Attack = 63-130, Word Identification = 71-148). The large range of word-level skills within groups produced correlations between these skills and reading comprehension (Word Attack = .82, Word Identification = .87) at the end of second grade that were very similar to those we have obtained with large random samples of children (Wagner, et al., 1996). Compared to the overall range of word level skills within groups, however, the differences between groups in these skills was relatively small. The more intensive comprehension instruction provided to children in the EP group provided them with language processing skills that were sufficient to compensate for the relatively small difference in word reading ability between them and children in the PASP group (Stanovich, 1984). Viewed the other way around, one could also say that the word level skills of children in the PASP condition were sufficiently stronger to aid them in constructing passage meaning as well as children who had received more than *twice* as much instruction in comprehension skills. The clear implication of this analysis is that maximally effective intervention programs for phonemically at-risk children should allow enough time for explicit and systematic instruction in *both* word level and comprehension oriented skills (cf. , Foorman, et al., 1998).

Although we are convinced from this study that effective intervention programs for children with serious phonological weaknesses must contain explicit and intense instruction in word-level skills, it is not

possible to identify which of these variables (explicitness or intensity) was most important to the word-level reading growth of children in the PASP (*essentially the program used by CRC*) condition. Because this condition provided both more explicit *and* more intensive in instruction in these skills, we cannot say whether its stronger effects were due to the specific nature of the instruction or to the fact that more time was spent on instruction and practice in this area. The time by activity analysis showed that 74% of the instructional activities in the PASP condition were spent teaching phonemic awareness and decoding skills (vs. 26% for the EP condition), so its effects may have resulted either from the amount of time spent, or the specific nature of the instruction. We would also like to point out, however, that if more explicit instruction is to cover the same content as less explicit instruction, it must *necessarily* require more time, because more specific elements, or skills, must be carefully taught.

In addition to comparisons among groups, we were also interested in the absolute level of reading growth obtained by children in our study. The average word level reading skills of children in the PASP (*essentially the program used by CRC*) condition were very close to the 50th percentile at the end of second grade. Passage Comprehension skills for children in both the PASP and EP groups were at the low end of the normal range, but they were entirely consistent with the overall estimated verbal ability of these children.

The average level of performance for both accuracy and fluency of phonemic reading skills in the PASP (*essentially the program used by CRC*) group is particularly striking for two reasons. First, there have been numerous observations of the difficulties of teaching these skills to children with phonological disabilities (Lovett, et al., 1994), and second, we have been much less successful in building fluency in these skills with older populations of disabled readers (Rashotte, Torgesen, & Wagner, 1997). What these results suggest to us is that, given the right instructional conditions, it is quite possible to help even children with core phonological weaknesses acquire the critical foundational skills in phonemic analysis and decoding that have been so strongly associated with good reading outcomes in elementary school.

Although the group scores on word level reading skills for children in the PASP (*essentially the program used by CRC*) condition fell within the average range, we also noted that there were significant individual differences within all our groups in response to instruction. Our analysis of the child variables associated with individual differences in response to the reading interventions suggests that the characteristics of "difficult-to-teach" children actually extend substantially beyond the domain of phonological weaknesses. Two of the most reliable predictors of response to intervention were the child's home environment (as measured by parental occupation and education) and rated level of behavior problems (attention, activity level) in the classroom. Even in the one-to-one teaching situation used in this study, the attention and behavior control problems of many of the children made it very difficult for them to profit from the instruction being provided to them. Since the children in this study received the significant bulk of their reading instruction in the regular classroom, this variable may also have been important because it was related to children's ability to profit from whole class instruction. This finding suggests that when we compare effectiveness across intervention studies conducted with different populations, we need to characterize the samples more broadly than simply in terms of their level of entering phonological skills.

As expected, the phonological variables (phonological awareness, rapid naming, and memory) also contributed uniquely to the prediction of individual differences in the growth of word-level reading skills. However, the pattern of results that emerged in this study was slightly different from that found when predictive relationships are studied in large random samples of children. In this latter type of study, phonological awareness and rapid naming have consistently emerged as significant unique predictors of growth in word reading ability (Wagner, et al., 1997), with the influences of phonological memory being redundant with the other two phonological abilities. In addition, in our research (Wagner, et al., 1994; Wagner, et al., 1997) phonological awareness has been a slightly stronger predictor of early growth in word reading ability than is rapid naming. In contrast, in the present study the most consistent

unique predictor of reading growth was rapid naming, followed by phonological memory and phonological awareness. This alteration of normal predictive relationships is almost certainly due to the restriction in range of beginning levels of phonological awareness within our sample. We only included children who performed within the bottom 12% on a measure of phonological awareness, while no criteria for level of rapid color and object naming, or for phonological memory, was used in the selection process. Clearly, these latter variables also suffered from restriction in range (since they are reliably correlated with phonological awareness), but because they were not selection variables themselves, the issue for them was less serious.

We also found, as have others (Vellutino, et al. 1996), that once phonological and home background factors are entered into the predictive equation, general verbal ability does not make a further independent contribution to understanding growth in word level reading skills. Even when the home background variable was not in the equation, general verbal ability did not make a significant unique contribution to the growth of word level reading skills beyond that captured by the phonological variables. This study provides a particularly strong test of role of general verbal ability in predicting growth of word level reading skills because children with a large range of verbal ability were included in the study. For example, at pretest, the estimated verbal IQ of children in the sample was 92.5 (range 76-126), and in the middle of second grade, the verbal IQ estimated from two subtests of the WISC-R was 89.2 (range 57-130). This finding provides clear support for the proposal (Lyon, 1995; Fletcher, et al., 1994) that level of general verbal ability may not be relevant in the early identification of children in need of special help in the development of word level reading skills. Although level of general verbal ability is not a unique predictor of response to instruction in word level skills, it did play a unique role in predicting growth in reading comprehension. These findings are consistent with a general view of reading growth in which phonological abilities provide the primary cognitive support for acquisition of word level skills, and both phonological abilities (because of their influence on word level reading) and general language abilities provide support for growth in reading comprehension.

Any large scale intervention study such as the one described in this paper is as much a study of implementation effectiveness as it is of specific instructional elements. Although our time by activity analyses, as well as our bi-weekly review of instructional videotapes, suggested a high degree of treatment fidelity across conditions, this does not mean that the instruction was delivered in the most effective manner possible. Although the interventions were delivered well enough to provide a fair test of their relative effectiveness against one another, it is important to acknowledge weaknesses that may have limited their absolute level of effectiveness, particularly for children with the most severe behavioral, home background, and cognitive weaknesses.

One virtue of this study is that we randomly assigned tutors to each of the two experimental interventions, but this virtue has a weakness associated with it. Prior to the initial training, none of the tutors had experience with the instructional procedures they were asked to implement. Thus, they began the study as relative instructional novices. Although our tutor's skills improved enormously over the course of the study, both of our expert instructional consultants identified weaknesses in implementation that occurred particularly at the beginning, but that were still not completely resolved at the end of the instruction.

Two other factors placed important limitations on the ultimate level of reading achievement of children in the study. Although it made our study economically feasible, using aides to alternate sessions with tutors was not an effective strategy with children as difficult to teach as the children in this study. Particularly for the PASP (*essentially the program used by CRC*) condition, in which much of the early instruction required use of a sophisticated questioning strategy, the aides could not be trained to a sufficiently high degree of skill to provide optimal kinds of practice and reinforcing activities. The level of instructional skill of the aides was more consistent with the requirements of the EP and RCS

conditions, in which straightforward practice tasks could be utilized that were more consistent with the core instructional procedures of the conditions.

One other important factor that most likely constrained the overall outcomes of this study was inconsistency between the instruction received in the one-to-one tutorials and that received in the regular classroom. In fact, the children may have sometimes received counter instructions in the two settings. For example, the PASP (*essentially the program used by CRC*) condition placed a premium on teaching children to use phonetic cues to decode novel words, but many of the classroom settings emphasized use of context as the first choice strategy. Furthermore, although most classroom teachers indicated that they taught "some phonics" in their reading curricula, very few employed what is now termed "systematic, explicit" phonics instruction as part of a completely balanced reading program in first and second grade. One recent large-scale intervention study (Foorman, et al., 1998) showed that reading instruction in the regular classroom containing systematic and explicit instruction in phonetic decoding skills is particularly beneficial for children who enter first grade low in phonological processing ability. Since the RCS condition was not the most effective instructional condition in the study, simply being consistent with classroom instruction is not an overriding virtue by itself. Rather, what may be required for many children like those included in this study is much more than 20 minutes a day of systematic, explicit instruction in phonemic decoding skills along with high quality comprehension oriented instruction and experiences. Thus, one recommendation for future research on tutorial or small group methods to prevent reading disabilities is that it should be conducted within the context of a regular classroom curriculum in reading that incorporates the kind of "balance" between word level and comprehension skills that is currently recommended by most experts in the field (Snow, Burns, & Griffin, 1998).

References

- Adams, M.J.(1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Agronin, M.E., Holahan, J.M. Shaywitz, B.A. & Shaywitz, S.E. (1992). The Multi-Grade Inventory for Teachers. In S.E. Shaywitz & B.A. Shaywitz (Ed.) *Attention deficit disorder comes of age* (pp. 29-67). Austin, TX: PRO-ED.
- Allington, R.L., & McGill-Franzen, A. (1994). Reading and the mildly handicapped. *International Encyclopedia of Education*, Oxford, U.K.: Pergamon.
- Bowey, J.A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, 87, 476-487.
- Brown, I.S. & Felton, R.H. (1990) Effects of instruction on beginning reading skills in children at risk for reading disability. *Reading and Writing: An Interdisciplinary Journal*, 2, 223-241.
- Ehri, L.C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. Metsala & L. Ehri (Eds.). *Word recognition in beginning reading*. (pp. 3-40). Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Fletcher, J.M., Shaywitz, S.E., Shankweiler, D.P., Katz, L., Liberman, I.Y., Stuebing, 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.

- Foorman, B.R., Francis, D.J., Fletcher, J.M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37-55.
- Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A., & Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*, 3-17.
- Fry, E. B., Kress, J.E. & Fountoukidis, D. L. (1993). *The reading teacher's book of lists: Third edition*. Englewood Cliffs, NJ.: Prentice Hall
- Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia, 46*, 3-20.
- Hannah, S. (1993). *Early Literacy Series, I & II*. Edmonton, Alberta CA: The Centre for Literacy.
- Hiebert, E. H., Colt, J. M., Catto, S. L., & Gury, E. C. (1992). Reading and writing of first-grade students in a restructured Chapter 1 program. *American Educational Research Journal, 29*, 545-572.
- Hollingshead, A.B. (1975). *Four factor index of social status*. New Haven, CN: Yale University, Department of Sociology.
- Iversen, S., & Tunmer, W. E. (1993) Phonological processing skills and the reading recovery program. *Journal of Educational Psychology, 85*, 112-126.
- Jastak, J. F., & Jastak, S. (1978). *The Wide Range Achievement Test- Revised*. Wilmington, Del.: Jastak Associates.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- Learning First Alliance (1998). *Every child reading: An action plan of the learning first alliance*. Washington, D.C.
- Lieberman, I.Y., Shankweiler, D., & Liberman, A.M. (1989). The alphabetic principle and learning to read. In Shankweiler, D. & Liberman, I.Y. (Eds.), *Phonology and reading disability: Solving the reading puzzle*(pp.1-33). Ann Arbor, MI: U. of Michigan Press.
- Lindamood, C. H., & Lindamood, P. C. (1984). *Auditory Discrimination in Depth*. Austin, TX: PRO-ED, Inc.
- Lovett, M.W., Borden, S.L., Deluca, T., Lacerenza, L., Benson, N.J., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. *Developmental Psychology, 30*, 805-822.
- Lyon, G.R. (1995). Towards a definition of dyslexia. *Annals of Dyslexia, 45*, 3-27.
- Mastropieri, M.A. & Scruggs, T.E. (1997). Best practices in promoting reading comprehension in students with learning disabilities: 1976-1996. *Remedial and Special Education, 18*, 197-213.

Metsala, J. & Ehri, L. C. (1998). *Word recognition in beginning reading*. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Pikulski, J. J. (1994). Preventing reading failure: A review of five effective programs. *The Reading Teacher*, 48, 30-39.

Pinnell, G.S., Lyons, C.A., Deford, D.E., Bryk, A., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 8-39.

Rashotte, C.A., Torgesen, J.K., & Wagner, R.K. (1997). *Differences in the growth of word reading accuracy and fluency after intensive instruction with older children*. Paper presented at the Annual meetings of the International Association for Dyslexia-Florida Branch. Miami, FL, September.

SAS Institute. (1997). *SAS/STAT software: Changes and enhancements through release 6.12*. Cary, NC: Author

Semel, E., Wiig, E. H., & Secord, W. (1987). *Clinical Evaluation of Language Fundamentals -- Revised*. Austin, TX: The Psychological Corporation

Shanahan, T. & Barr, R. (1995). Reading recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 30, 958-996.

Share, D.L., & Stanovich, K.E. (1995). Cognitive processes in early reading development: A model of acquisition and individual differences. *Issues in Education: Contributions from Educational Psychology*, 1, 1-57.

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

Slavin, R.E., Madden, N.A., Karweit, N.L., Libvermon, B.J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for referring urban education. *American Educational Research Journal*, 27, 255-278.

Smith, L. (1992). *Auditory discrimination reading series-sequence B*. Independence, MO.: Poppin Creations.

Snow, C.E., Burns, M.S. & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stanovich, K. (1984). The interactive-compensatory model of reading: a confluence of developmental, experimental, and educational psychology. *Remedial and Special Education*, 5, 11-19.

Tangel, D.M. & Blachman, B.A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior*, 24, 233-261.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *Guide for administering and scoring, the Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago: Riverside Publishing.

- Torgesen, J.K. (in press). Phonologically based reading disabilities: Toward an integrated theory of one kind of learning disability. In L. Spear-Swerling & R.J. Sternberg (Eds.), *Perspectives on learning disabilities*. Mahwah, NJ: Erlbaum Associates
- Torgesen, J.K., & Burgess, S.R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala & L. Ehri (Eds.). *Word Recognition in Beginning Reading*. (pp. 161-188). Hillsdale, NJ: Lawrence Erlbaum Assoc.
- Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1997). The prevention and remediation of severe reading disabilities: keeping the end in mind. *Scientific Studies of Reading*, 1, 217-234.
- Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.
- Torgesen, J.K., Wagner, R.K., Rashotte, C.A., Alexander, A.W., & Conway, T. (1997). Preventive and remedial interventions for children with severe reading disabilities. *Learning Disabilities: An Interdisciplinary Journal*, 8, 51-62.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small S. G., Pratt, A., Chen R., & Denckla, M. B. (1996). 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. *Journal of Educational Psychology*, 88, 601-638.
- Wagner, R.K., Torgesen, J.K., Laughon, P., Simmons, K., & Rashotte, C.A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology*, 85, 1-20.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C.A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology*, 30, 73-87.
- Wagner, R.K., Torgesen, J.K., Rashotte, C.A., Hecht, S.A., Barker, T.A., Burgess, S.R., Donahue, J., & Garon, T. (1997). Changing causal relations between phonological processing abilities and word-level reading as children develop from beginning to fluent readers: A five-year longitudinal study. *Developmental Psychology*, 33, 468-479.
- Wasik, B.A., & Slavin, R.E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 179-199.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children: Revised*. New York: The Psychological Corporation.
- Wiederholt, J.L. & Bryant, B.R. (1992). *Gray Oral Reading Tests--III*. Austin, TX: PRO-ED.
- Wolf, M. (1997). A provisional, integrative account of phonological and naming-speed deficits in dyslexia: Implications for diagnosis and intervention. In B. Blachman (Ed.), *Foundations of Reading Acquisition and Dyslexia: Implications for Early Intervention* (pp.67-91). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Woodcock, R.W. (1987). *Woodcock Reading Mastery Tests-Revised*. Circle Pines, Minn.: American Guidance Service.

Woodcock, R.W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM/Teaching Resources