LINNEA C. EHRI and SIMONE R. NUNES, City University of New York Graduate Center; STEVEN A. STAHL, University of Georgia; DALE M. WILLOWS, Ontario Institute for Studies in Education, University of Toronto

Systematic Phonics Instruction Helps Students Learn to Read: Evidence From the National Reading Panel's Meta-Analysis

Abstract: A quantitative meta-analysis evaluating the effects of systematic phonics instruction compared to unsystematic or no-phonics instruction on learning to read was conducted using 66 treatment–control comparisons derived from 38 experiments. The overall effect of phonics instruction on reading was moderate, *d* = 0.41. Effects persisted after instruction ended. Effects were larger when phonics instruction began early (*d* = 0.55) than after first grade (*d* = 0.27). Phonics benefited decoding, word reading, text comprehension, and spelling in many readers. Phonics helped low and middle SES readers, younger students at risk for reading disability (RD), and older students with RD, but it did not help low achieving readers that included students with cognitive limitations. Synthetic phonics and larger-unit systematic phonics programs produced a similar advantage in reading. Delivering instruction to small groups and classes was not less effective than tutoring. Systematic phonics instruction helped children learn to read better than all forms of control group instruction, including whole language. In sum, systematic phonics instruction proved effective and should be implemented as part of literacy programs to teach beginning reading as well as to prevent and remediate reading difficulties.

In 1997, the U.S. Congress directed that a national panel be convened to review and evaluate research on the effectiveness of various approaches for teaching children to read. The Director of the National Institute of Child Health and Human Development (NICHD) in consultation with the Secretary of Education constituted the National Reading Panel (NRP) composed of 14 individuals. Members of the Panel formed subgroups to review research on alphabetics, comprehension, fluency, teacher education, and technology. The alphabetics subgroup conducted two meta-analyses, one on phonemic awareness instruction (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001) and another on systematic phonics instruction. Results of the phonics meta-analysis are presented in this article. The final report incorporating all the subgroup reviews was submitted to Congress in April 2000 (NRP, 2000).

Teaching students to read is a complex task. Children enter school with substantial competence speaking their language, but typically they have little knowledge about how to read and write. The purpose of literacy instruction in schools is to help children master the many challenges of written language. These include acquiring knowledge of the alphabetic system,

Journal of Direct Instruction Vol. 2, No. 2, pp. 121–166.

Review of Educational Research, Fall 2001, Vol. 71(3), *393–447.*

learning to decode new words, building a vocabulary of words that can be read from memory by sight, and becoming facile at constructing, integrating, and remembering meanings represented in text. In order for children to be able to link their knowledge of spoken language to their knowledge of written language, they must master the alphabetic code, that is, the system of grapheme–phoneme correspondences that links the spellings of words to their pronunciations. Phonics instruction teaches beginning readers the alphabetic code and how to use this knowledge to read words. In systematic phonics programs, a planned set of phonics elements is taught sequentially. The set includes not only the major correspondences between consonant letters and sounds but also short and long vowel letters and sounds, and vowel and consonant digraphs (e.g., oi, ea, sh, th). It also may include blends of letter–sounds that form larger subunits in words. The larger units taught might include onsets (consonants that precede vowels, such as "j" in *jump* or "st" in *stop*) and rimes (i.e., the vowel and following consonants such as "ump" in *jump* and "op" in *stop*).

Over the years, educators have disagreed about how beginning reading should be taught. Some have advocated starting with a systematic phonics approach whereas others have argued for whole-word or whole-language approaches. Disagreement has centered on whether teaching should begin with explicit instruction in symbol–sound correspondences or with whole words, or whether initial instruction should be meaning-centered with correspondences taught incidentally in context as needed. The current view is that, because research suggests that systematic phonics approaches are more effective than nonsystematic approaches, children should be provided with systematic phonics instruction as part of a balanced reading program.

The purpose of this review was to examine the research evidence to determine whether systematic phonics instruction helps children learn to read more effectively than unsystematic phonics instruction or instruction teaching little or no phonics. Is phonics instruction more effective under some circumstances than others, such as tutoring versus small groups or classrooms; beginning grades as opposed to later grades; for children who are progressing normally in reading as well as for children who are at risk or disabled in their reading? Does phonics instruction improve children's reading comprehension as well as their word-reading and spelling skills? Does the type of instruction given to control groups to evaluate the effectiveness of phonics instruction (e.g., whole-word or whole-language approaches) make a difference?

A meta-analysis was conducted to evaluate the evidence. The Panel searched the literature to locate experimental studies that administered systematic phonics instruction to one group of children and administered another type of instruction involving either unsystematic phonics or no phonics to a control group. The review was limited to experiments because these provide the strongest evidence that instruction rather than some other factor caused the improvement in reading. The studies had to examine phonics programs of the sort used in schools rather than used in laboratory experiments focused on single processes. The studies had to measure reading as an outcome of instruction. Studies were excluded if they were in the Panel's other meta-analysis examining effects of phonemic awareness instruction (Ehri et al., 2001). To insure that they met the research standards of the field, studies were limited to those published in peer reviewed journals. A total of 38 studies was identified and coded for various characteristics of students, instruction, and experimental design. The meta-analysis examined the size of effects that resulted when treatment and control groups were compared on reading and spelling outcomes.

Our main interest for this review was in reading as an outcome of instruction. Reading may be defined in various ways. Although many people consider the term reading to refer to text comprehension, the meaning is broader and encompasses reading words as well, for

example, on grocery lists, in telephone books, on labels and signs, and on computer screens. Reading pseudowords is also considered reading, because this task assesses the ability to pronounce unknown written words. In their *Literacy Dictionary*, Harris & Hodges (1995) identify 13 different definitions of reading. In their view, the variety "amply demonstrate(s) that such definitions need to be seen in the context of the theoretical and pragmatic orientations of the definer" (pp. 206–207). Our approach was pragmatic, reflecting the various ways that researchers measured reading outcomes in their studies. Reading included reading real words and pseudowords, reading text orally, and text comprehension.

Phonics Instruction

Although they are often confused, phonics instruction is different from phonemic awareness instruction. The goal of phonemic awareness (PA) instruction is to teach children to focus on and manipulate phonemes in spoken words, for example, blending sounds to form words (/t/-/o/-/d/ = "toad"), or segmenting words into phonemes ("shock" = /š/-/a/-/k/). Some PA programs teach children to use letters to manipulate phonemes in speech. This makes them more similar to phonics programs that may teach children to sound out and blend letters to decode words or to segment words into phonemes to spell words. However, phonics programs typically cover more than this and include instruction and practice in reading words in and out of text.

Several different approaches have been used to teach phonics systematically (Aukerman, 1971, 1984; Harris & Hodges, 1995). These include synthetic phonics, analytic phonics, embedded phonics, analogy phonics, onset–rime phonics, and phonics through spelling. These approaches differ in several respects. Synthetic phonics programs use a part-to-whole approach that teaches children to convert graphemes into phonemes (e.g., to pronounce each letter in *stop*, /s/-/t/-/a/-/p/) and then to blend the phonemes into a recognizable word. Analytic phonics uses a whole-to-part approach that avoids having children pronounce sounds in isolation to figure out words. Rather children are taught to analyze letter–sound relations once the word is identified. For example, a teacher might write the letter P followed by several words, *put*, *pig*, *play*, *pet*. She would help students read the words and recognize that they all begin with the same sound that is associated with P. Phonics-through-spelling programs teach children to segment and write the phonemes in words. Phonics in context teaches children to use letter–sound correspondences along with context cues to identify unfamiliar words they encounter in text. Analogy phonics teaches children to use parts of written words they already know to identify new words. For example, they are taught a set of key words that are posted on the wall (e.g., *tent*, *make*, *pig*) and then are taught to use these words to decode unfamiliar words by segmenting the shared rime and blending it with the new onset (e.g., *rent, bake*, *jig*). Some systematic phonics programs are hybrids that include components of two or more of these approaches.

Phonics programs may differ in several other important ways, for example,

- how many letter–sound relations are taught and how they are sequenced;
- whether phonics generalizations or phonemic awareness is taught and at what pace;
- whether learning activities include oral drill-and-practice or reciting phonics rules or filling out worksheets;
- whether children read decodable text in which the vocabulary is limited mainly to words containing familiar letter–sound associations;
- whether phonics instruction is embedded in or segregated from the literacy curriculum;
- whether the teaching approach involves direct instruction in which the teacher takes an active role and students passively

respond, or a "constructivist" problem-solving approach is used, and

- how interesting and motivating the instructional activities are for teachers and for students (Adams, 1990; Aukerman, 1981).

The phonics programs examined for this review varied in many of these ways, so phonics was not taught uniformly across programs. We had hoped to examine whether these properties influenced the programs' effectiveness, but we found that many studies did not provide sufficient detail to code the studies for these properties. The only property that we investigated was whether programs emphasized a synthetic approach or whether the emphasis was on larger subunits of words.

Synthetic phonics programs teach children systematically and sequentially the correspondences between graphemes and phonemes of the language and how to apply them to decode unfamiliar words by sounding out the letters and blending them. One potential difficulty in blending sounds is that children must learn to delete the "extra" schwa vowel sounds produced when letters are pronounced separately. For example, blending "tuh-a-puh" requires deleting the "uh" sounds to produce the blend "tap." Another difficulty is that children must remember all the sounds in the proper order to blend them correctly.

Larger-unit programs are thought to ease these difficulties. Children are taught to decode subunits such as ST, AP, EAM, as chunks, thus reducing the number of word parts to sound out and blend and reducing the schwa problem. Teaching children to pronounce parts of words provides the basis for teaching them the strategy of reading new words by analogy to known words (e.g., reading *stump* by analogy to *jump*).

Evaluating the effectiveness of systematic phonics instruction has been addressed many times in the literature. The best known effort was Jeanne Chall's (1967) comprehensive review of beginning reading instruction covering studies up to the mid-1960s, *Learning to Read: The Great Debate*. Her basic finding was that early and systematic instruction in phonics led to better achievement in reading than later and less systematic phonics instruction. In the 1967 edition of her review, Chall did not recommend any particular type of phonics instruction, but in the 1983 edition she suggested that synthetic phonics instruction held a slight edge over analytic phonics instruction. Chall's basic finding has been reaffirmed in many research reviews conducted since then (e.g., Adams, 1990; Anderson, Hiebert, Wilkinson, & Scott, 1985; Balmuth, 1982; Dykstra, 1968).

Alternative Nonsystematic Phonics Instruction

At the time of Chall's (1967) original review, the contrast between phonics instruction and the alternative "look–say" methods was considerable. In the look–say approach, children were taught to read words as wholes, and they practiced reading words until they had acquired perhaps 50 to 100 in their sight vocabularies. Only after this, toward the end of first grade, did phonics instruction begin. This was truly a nonphonics approach, because teaching letter–sound relations was delayed for a considerable time.

More recently, whole language approaches have replaced the whole word method as the most common alternative to systematic phonics programs. The shift has involved a change from very little letter–sound instruction to a modicum of letter–sounds taught unsystematically in first grade. Whole language teachers are not told to wait until a certain point before teaching children about letter–sound relationships. Typically they provide some instruction in phonics, usually as part of invented spelling activities or through the use of graphophonemic prompts during reading (Routman, 1996). However, their approach is

to teach it unsystematically in context as the need arises. Observations suggest that in whole-language classrooms, instruction in vowel letter–sound correspondences occurs infrequently (Stahl, Duffy-Hester, & Stahl, 1998). This contrasts with systematic phonics programs where the teaching of vowels is central in learning to decode (Shankweiler & Liberman, 1972).

In the present meta-analysis, the effectiveness of systematic phonics instruction was compared to various types of nonphonics or unsystematic phonics instruction given to control groups. Some studies provided whole language instruction or whole word instruction to control groups. Another form of control-group instruction involved some type of basal program. In basal programs, teachers are provided with a structured package of books and supplementary materials. They work from a manual that details daily lesson plans based on a scope and sequence of the reading skills to be taught. Students are given workbooks to practice skills. Tests are used to place students in the proper levels of the program and to assess mastery of skills (Aukerman, 1981). Basal reading programs of the same era tend to be roughly similar in their characteristics. The basal programs given to control groups in our studies provided only limited or no systematic phonics instruction. Typically they were the programs prescribed in a school or district. A few studies created control groups by using the performance of comparable classes of students enrolled in the same schools the year prior to the treatment when phonics was not taught systematically (Snider, 1990; Vickery, Reynolds, & Cochran, 1987). Some studies included more than one control group. Selected for the calculation of effect sizes in the meta-analysis was the group receiving the least phonics instruction. In the text below, we have referred to the control treatments in various ways, as unsystematic or nonsystematic phonics or no phonics. These terms are meant to refer to the entire pool of control treatments and should be regarded as synonymous.

Delivery Systems for Teaching Phonics

One-on-one tutoring is the preferred form of instruction for students who are having difficulties, because lessons can be tailored to individual needs. Eight studies examined phonics taught by tutoring. In the remaining studies, phonics instruction was delivered to small groups or whole classes. We examined whether one type of delivery system enhanced reading more than the other types. In the NRP meta-analysis of phonemic awareness instruction (Ehri et al., 2001), small groups were found to produce statistically larger effect sizes than tutoring or whole classrooms.

Grade and Reading Ability

A question of particular interest in the field is when should phonics instruction begin? It has been suggested (Chall, 1996b) that beginners need to develop foundational knowledge such as concepts about print, phonological awareness, and letter names prior to formal reading instruction. Expecting students to grapple with synthetic phonics and decoding instruction in kindergarten may be too much. On the other hand, countries such as New Zealand and the United Kingdom have introduced children to reading and writing at the age of 5 years in full-day programs for many years. Thus, the notion that kindergartners are not ready for formal reading instruction at age 5 is questionable.

Some studies in our review introduced kindergartners to simplified reading and spelling activities. Instruction began by providing a foundation for students and then building on this to ease students into reading when they became ready (Blachman et al., 1999; Vandervelden & Siegel, 1997).

In the United States, formal reading instruction typically begins in first grade, so introducing phonics instruction above first grade means that students have already acquired some reading ability presumably from another

method. To exert an impact at this point may be harder because it may require students to change their way of processing print. Our review included studies that introduced phonics to students from kindergarten to sixth grades. Of interest was whether phonics instruction was more effective in kindergarten and first grades than in later grades.

Phonics instruction is considered particularly beneficial to children with reading problems. Studies indicate that students with reading disability (RD) have exceptional difficulty decoding words (Rack, Snowling, & Olson, 1992). Phonics instruction that teaches decoding is expected to remediate this deficit and enable these students to make better progress in learning to read. Of interest in our meta-analysis was whether phonics instruction helps to prevent reading failure among beginners showing signs of being at risk, and whether it helps to remediate reading difficulties among older struggling readers.

Two types of children with reading problems have been distinguished by researchers, children whose reading level falls well below their cognitive abilities making them unexpectedly poor readers, and children whose poor reading is not surprising given that their cognitive abilities are below average as well. Various labels such as dyslexic or learning disabled or reading disabled have been applied to children showing a discrepancy between IQ and reading level. Some studies in our review were conducted with unexpectedly poor readers while other studies were conducted with poor readers whose cognitive abilities either were not assessed or were found to be below average. We examined whether phonics instruction helped to remediate reading difficulties separately in the two cases, that is, in reading-disabled students and in low achieving readers.

Word Reading Outcomes

Readers use various processes to read words (Ehri, 1991, 1994), and all of these processes benefit from alphabetic knowledge which is the goal of phonics instruction. Decoding words involves converting graphemes into phonemes and blending them to form recognizable words, or blending larger subunits into words. Synthetic phonics programs teach children the grapheme–phoneme blending routine explicitly. Reading words by analogy involves using parts of known words to read new words having the same parts. Larger-unit phonics programs teach this routine explicitly. Reading words by sight involves retrieving from memory words that the reader has already learned to read. To remember how to read sight words, knowledge of the alphabetic system is required to establish connections between spellings of words and their pronunciations in memory (Berninger et al., 2001; Ehri, 1992, 1998; Perfetti, 1992). To predict unfamiliar words in text, readers may use both letter–sound cues and context cues (Tunmer & Chapman, 1998).

One purpose of the meta-analysis was to determine whether there is evidence that phonics instruction improves readers' ability to read words in various ways. Among the studies examined were those in which the ability to decode words was tested by giving children regularly spelled words and pseudowords to read. Sight vocabulary was examined by having children read miscellaneous words that included irregularly spelled words ordered by grade level. In addition to word reading outcomes, measures of reading fluency, comprehension, and spelling were also investigated in the meta-analysis.

Method

Database

An electronic search was conducted in two databases, ERIC and PsychInfo. Three sets of terms derived from various reference guides were used to locate all articles indexed by these terms (Barr, Kamil, Mosenthal, & Pearson, 1991; Flood, Jensen, Lapp, & Squire, 1991; Harris & Hodges, 1995; Pearson, Barr, Kamil, & Mosenthal, 1984; Purves, 1994):

Set 1: alphabetic code, analogy approach, code emphasis, compare–contrast, decodable text, decoding, phonemic decoding, phonetic decoding, phonological decoding, direct code, direct instruction, *Reading Mastery*, explicit instruction, explicit phonological processes, grapheme–phoneme correspondences, grapho-phonic, Initial Teaching Alphabet, letter training, letter–sound correspondences, linguistic method, McCracken, Orton-Gillingham, phoneme analysis, phoneme blending, phoneme–grapheme correspondences, phonics, alphabetic phonics, analytic phonics, embedded phonics, structured phonics, synthetic phonics, systematic phonics, phonological processing, Recipe for Reading, recoding, phonological recoding, Slingerland approach, Spaulding approach, spelling, word study, word sort, words by analogy.

Set 2: beginning reading, beginning reading instruction, instruction, intervention, learning to decode, reading improvement, reading instruction, remedial training, remedial reading, remediation, teaching, training, disabled readers, dyslexia, reading difficulties, reading disability, reading failure, reading problems.

Set 3: miscues, oral reading, reading ability, reading achievement, reading acquisition, reading aloud, reading comprehension, reading development, reading processes, reading skills, silent reading, story reading, word attack, word identification, word recognition, word reading, nonword reading.

The search uncovered 643 articles in PsychInfo and 730 articles in ERIC.

The following criteria were applied to screen studies for the analysis. Studies had to:

- adopt an experimental or quasi-experimental design with a control group.
- appear in a refereed journal. Peer review served to insure that the quality of the studies met research standards in the field. This criterion was adopted and applied to all the reviews conducted by the NRP.
- be published after 1970. Limiting the time period yielded a manageable and contemporary database.
- focus on the teaching of phonics in English and be published in English.
- compare the effectiveness of instruction in systematic phonics with that of instruction providing unsystematic phonics or no phonics instruction.
- measure reading as an outcome.
- report statistics permitting the calculation or estimation of effective sizes.
- involve interventions that might be found in schools. Short-term laboratory studies and studies that involved teaching of very limited alphabetic processes were excluded.

Studies did not include those already in the National Reading Panel's meta-analysis of phonemic awareness training studies (Ehri et al., 2001). This allowed independent conclusions to be drawn about the effectiveness of these two forms of instruction.[1]

Abstracts of the studies that we located in the electronic search were reviewed to select those appearing to meet our criteria. Seventy-one were located and identified for more complete inspection. Most of the studies were rejected because they did not examine systematic phonics instruction or were not experiments or quasi-experiments with a control group. Few if any experiments were rejected because they were conducted prior to 1970. We eliminated short-term laboratory studies with a limited focus (14 studies), studies lacking a nonsystematic phonics control group (5 studies), studies lacking reading as an outcome or lacking any statistics allowing the calculation of effect size (11 studies), duplicate studies reporting the same data (5 studies), and studies that did not examine phonics instruction (2 studies). Remaining in the pool were 34 studies.

We deviated from our criteria and search procedure in four cases. First-year findings of a 3-year phonemic awareness plus phonics study by Blachman, Ball, Black, and Tangel (1994)

and Blachman et al. (1999) had been included in the NRP phonemic awareness meta-analysis (Ehri et al., 2001). Four studies were published or were in press after the electronic search and were forwarded to us by the authors (Blachman et al., 1999; Lovett et al., 2000; Stuart, 1999; Torgesen et al., 1999).

From the 38 studies entered into the database, 66 treatment–control group comparisons were derived, a number more than adequate for conducting a meta-analysis (Rosenthal, 1991). In six cases the same control group was compared to two different phonics treatment groups. In one study the same control group was compared to four different treatments (Lovett et al., 2000). Although this meant that effect sizes were not completely independent across comparisons, we preferred this alternative to that of combining treatment and control groups with studies, because we did not want to obscure important moderator variables of interest.

Studies were coded for several characteristics that were included as moderators in the meta-analysis:

- Type of phonics program (synthetic, larger subunits, miscellaneous);
- Specific phonics program if replicated in at least three comparisons (see descriptions of programs in Appendix A);
- Type of control group (basal, regular instruction, whole language, whole word, miscellaneous);
- Sample size;
- Grade level or age (kindergarten, first, second through sixth);
- Reading ability (normally achieving, at risk, reading disabled, low achieving);
- Socioeconomic status (low, middle, varied, not given);
- Instructional delivery unit (class, small groups, tutoring);
- Group assignment procedure (random assignment, nonequivalent groups);
- Existence of pretreatment group differences (present, absent, present but posttest means adjusted, not given).

The studies, their properties, and effect sizes are listed in Appendix B.

The length of treatment was not used as a moderator variable. Many of the studies were vague about the amount of time devoted to phonics instruction, so calculating time precisely was not possible, particularly in classroom studies which provided instruction regularly throughout the school year.

The students participating in the studies were categorized by type of reader:

1. Normally achieving (NA) readers: children who either were not screened for reading ability or were screened to exclude poor readers;
2. At risk (AR) readers: kindergartners and first graders judged to be at risk for future reading difficulties because of poor letter knowledge, poor phonemic awareness, poor reading skills, or enrollment in low achieving schools;
3. Students with a reading disability (RD): children who were below grade level in reading but at least average cognitively and were above first grade in most cases;
4. Low achieving (LA) readers: children above first grade who were below average readers and whose cognitive level was below average or was not assessed.

Studies in the database were published between 1970 and 2000, although the majority were conducted in the last 10 years: 1970 to 1979 (1 study); 1980 to 1989 (9 studies); and 1990 to 2000 (28 studies). Most (66%) were carried out in the United States, but 24% were done in Canada, and the remainder in the United Kingdom, Australia, and New Zealand.

Performance on six specific outcomes assessing reading or spelling was analyzed: decoding regularly spelled real words; decoding pseudowords; reading real words that included irregularly spelled words; spelling words correctly or according to developmental criteria (Morris & Perney, 1984; Tangel & Blachman, 1995); comprehending text; and reading connected text orally. A few studies used general reading tests to assess outcomes. Tasks measuring reading or spelling of words that were taught directly during phonics instruction were not included as outcomes. Outcomes that did not fit into the above categories (e.g., letter–sound knowledge and attitudes) were not entered into the database.

In categorizing outcome measures, no distinction was drawn between standardized and experimenter-devised tests. The comprehension measures tended to be standardized. Oral reading measures tended to be informal reading inventories that were neither standardized nor developed specifically for the study. Word lists were both standardized and experimenter devised.

Performance of students was measured at various times before, during, and after instruction. Entered into the database were outcomes of posttests measured at three possible points in time: at the end of training; at the end of the first school year if the program was taught for more than 1 year; and after a delay to assess long-term effects of training. The most common posttest occurred at the end of treatment or at the end of one school year for treatments that continued longer than a year, so this assessment of reading was the outcome used in most of the analyses of moderator variables.

Meta-Analysis

The primary statistic used to analyze effects of phonics instruction on outcome measures was effect size, indicating whether and by how much performance of the treatment group exceeded performance of the control group, with the difference expressed in standard deviation units. The formula used to calculate raw effect sizes for each treatment–control comparison consisted of the mean of the treatment group minus the mean of the control group divided by a pooled standard deviation. This formula was adopted for use in all meta-analyses conducted by the NRP.

Some studies administered more than one task to measure specific outcomes. When this occurred, effect sizes were calculated separately on each measure and then averaged to create one effect size for that outcome. This step insured that no single treatment–control comparison contributed more than one effect size to any single outcome category in the meta-analysis.

For each of the 66 treatment–control comparisons, effect sizes across the six specific outcomes were averaged to create one overall effect size for that comparison. The overall effect size was interpreted to indicate the general impact of phonics instruction on learning to read. Although one of the six outcomes contributing to the overall average was a spelling measure, spelling effect sizes contributed only 16% of the values while reading measures contributed 84% to the average. Studies have shown that reading words and spelling words are highly correlated, with *r*s commonly above .70, indicating that both involve the same processes (Ehri, 1997). Thus, interpreting the overall effect size as an index of reading is justified.

The information required to generate and analyze effect sizes was entered into the database using Microsoft Excel and SPSS. The data included codes identifying each treatment–control comparison, codes characterizing properties of the comparisons, means and standard deviations of treatment and control groups on outcome measures, pooled standard deviations, and raw effect sizes (g). Because the formula for g overestimates population effect sizes to the extent that sample sizes are small, the formula was corrected to yield an unbiased estimator of the population effect size (d) (Johnson & Eagly, 2000). When

means and standard deviations were not available, the DSTAT program was used to estimate effect sizes based on *t* or *F* values (Johnson, 1989).

The means and standard deviations that were used to calculate effect sizes were verified by checking all of them at least twice. Intercoder reliability (i.e., percentage agreement) was conducted on the variables used in the meta-analysis and exceeded the NRP-prescribed level of 90%. Disagreements were resolved by discussion and consensus.

The DSTAT statistical package (Johnson, 1989) was employed to calculate mean effect sizes and to test the influence of moderator variables using a fixed-effects model. Statistical tests assessed whether mean weighted effect sizes (*d*) were significantly greater than zero, whether mean effect sizes were derived from a homogeneous set (*Q* statistic), and whether pairs of mean effect sizes differed significantly for different levels of a moderator variable ($p < .05$). The analysis did not include tests of interactions between moderator variables because the numbers of comparisons were insufficient in many cases.

Results

Meta-Analysis

The statistic used to assess the effectiveness of systematic phonics instruction was effect size which measures how much the mean of the phonics treatment group exceeded the mean of the control group in standard deviation units. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean while an effect size of 0 indicates that treatment and control group means were identical. According to Cohen (1988), an effect size of 0.20 is considered small, an effect size of 0.50 is moderate, and an effect size of 0.80 or above is large. Translated into percentiles, $d = 0.20$ indicates that the treatment has moved the average child from the 50th to the 58th percentile; $d = 0.50$ indicates that the treatment has moved the average child to the 69th percentile; $d = 0.80$ indicates that the treatment has moved the average child to the 79th percentile. The complete list of comparisons, codings, and effect sizes is given in Appendix B.

Effect sizes were calculated for each of the 66 treatment–control group comparisons to assess the impact of phonics instruction. These were analyzed using the DSTAT program (Johnson, 1989). Table 1 reports several statistics from this analysis including mean effect sizes weighted by sample size (*d*). Figure 1 displays several effect sizes. Effect sizes associated with each level of the moderator variables are given along with a symbol indicating whether values were statistically greater than zero and a 95% confidence interval. The number of comparisons contributing to each effect size is listed. Effect sizes based on larger numbers of comparisons are more reliable and representative of the population than effect sizes based on small numbers. In interpreting effect sizes, particularly those that are not statistically significant, it is important to note whether the number of studies was sufficient to yield acceptable statistical power.

Results of the *Q* model–fit statistical test of homogeneity are presented. This statistic assesses the plausibility that the underlying effects are gauging the same population effect. When *Q* is large and significant indicating that the pool of effect sizes is quite variable, the model fits poorly and effects should be interpreted conservatively. They may depend upon other moderator variables.

Table 1 reports the results of tests to compare pairs of effect sizes across levels of several moderator variables to determine whether one level was statistically greater than another. Those proving different are listed under Contrasts. Table 2 reports statistics summarizing unweighted effect sizes (*g*) including means, medians, standard deviations, and minimum and maximum values.

Table 1

*Mean Effect Sizes (*d*) as a Function of Moderator Variables With Effect Sizes Measured Immediately After Training or At The End of One School Year When Training Lasted Longer Unless Otherwise Labeled*

Moderator Variable and Level	Number Cases	Mean *d* [a]	Homogen.[b]	95% CI	Contrast[c]
Time of Posttest					
End of Training	65	0.41*	No	0.36 to 0.47	n.s.
End of Training or First Year	62	0.44*	No	0.38 to 0.50	
End of Training[d]	6	0.51*	Yes	0.32 to 0.70	n.s.
Follow-up[d]	6	0.27*	Yes	0.07 to 0.46	
Outcome Measures					
Decoding regular words (DRW)	30	0.67*	No	0.57 to 0.77	DRW = DP;
Decoding pseudowords (DP)	40	0.60*	No	0.52 to 0.67	Both >
Reading misc. words (RMW)	59	0.40*	No	0.34 to 0.46	RMW, SW,
Spelling words (SW)	37	0.35*	No	0.28 to 0.43	RTO, CT
Reading text orally (RTO)	16	0.25*	No	0.15 to 0.36	
Comprehending text (CT)	35	0.27*	No	0.19 to 0.36	
Characteristics of Participants					
Grade Levels					
Kindergarten & 1st	30	0.55*	No	0.47 to 0.62	Kind.–1st >
2nd–6th (NA, RD, LA[e])	32	0.27*	Yes	0.18 to 0.36	2nd–6th
Younger Grades					
Kindergarten	7	0.56*	Yes	0.40 to 0.73	n.s.
1st Grade	23	0.54*	No	0.46 to 0.63	
Grade and Reading Ability					
Kindergarten At Risk	6	0.58*	Yes	0.40 to 0.77	1AR > 2-6N,
1st Normal Achieving	14	0.48*	No	0.38 to 0.58	2-6LA, RD
1st At Risk (1AR)	9	0.74*	No	0.56 to 0.91	
2nd–6th Normal Ach. (2–6N)	7	0.27*	Yes	0.12 to 0.43	
2nd–6th Low Ach. (2–6LA)	8	0.15n.s.	Yes	-0.06 to 0.36	
Reading Disabled (RD)	17	0.32*	Yes	0.18 to 0.46	
Outcome Measures					
Kindergarten and First Graders					
Decod. Regular words (DRW)	8	0.98*	No	0.81 to 1.16	DRW > RMW,
Decoding pseudowords (DP)	14	0.67*	No	0.56 to 0.78	CT, RTO;
Reading misc. words (RMW)	23	0.45*	No	0.37 to 0.53	SW > RTO;
Spelling words (SW)	13	0.67*	No	0.54 to 0.79	DP > RTO
Reading text orally (RTO)	6	0.23*	No	0.05 to 0.41	
Comprehending text (CT)	11	0.51*	No	0.36 to 0.65	
2nd–6th (NA, RD, LA[e])					
Decod. Regular words (DRW)	17	0.49*	No	0.34 to 0.65	DRW > SW;
Decoding pseudowords (DP)	13	0.52*	Yes	0.37 to 0.66	DP > SW, CT
Reading misc. words	23	0.33*	No	0.22 to 0.44	
Spelling words (SW)	13	0.09 n.s.	Yes	-0.04 to 0.23	
Reading text orally	6	0.24*	Yes	0.08 to 0.39	
Comprehending text (CT)	11	0.12 n.s.	Yes	-0.04 to 0.28	
Socioeconomic Status					
Low SES	6	0.66*	Yes	0.48 to 0.85	n.s.
Middle SES	10	0.44*	No	0.28 to 0.60	
Varied	14	0.37*	Yes	0.26 to 0.48	
Not Given	32	0.43*	No	0.34 to 0.51	

continued on next page

Table 1, continued

Moderator Variable and Level	Number Cases	Mean *d* [a]	Homogen.[b]	95% CI	Contrast[c]
Characteristics of Instruction					
Type of Phonics Program					
Synthetic	39	0.45*	No	0.39 to 0.52	n.s.
Larger Phon. Units[f]	11	0.34*	No	0.16 to 0.52	
Miscellaneous	10	0.27*	Yes	0.08 to 0.46	
Specific Phonics Programs[g]					
NRS-Beck LRDC (S)	4	0.47*	Yes	0.33 to 0.60	n.s.
Direct Instruction (S)	4	0.48*	No	0.13 to 0.83	
Lovett Direct Instruct (S)	4	0.41*	Yes	0.04 to 0.77	
Lovett Analogy (LU)	4	0.48*	Yes	0.11 to 0.86	
Lippincott (S)	3	0.68*	Yes	0.43 to 0.93	
Orton-Gillingham (S)	10	0.23*	Yes	0.06 to 0.39	
Sing Spell Read Write (S)	3	0.35*	Yes	0.21 to 0.50	
Instructional Delivery Unit					
Tutor[f]	8	0.57*	No	0.38 to 0.77	n.s.
Small Group	27	0.43*	Yes	0.34 to 0.52	
Class	27	0.39*	No	0.31 to 0.48	
Type of Control Group					
Basal	10	0.46*	Yes	0.37 to 0.55	n.s.
Regular Curriculum	16	0.41*	No	0.27 to 0.54	
Whole Language	12	0.31*	No	0.16 to 0.47	
Whole Word	10	0.51*	No	0.35 to 0.67	
Miscellaneous	14	0.46*	Yes	0.28 to 0.63	
Characteristics of the Design of Studies					
Assignment of Participants to Treatment and Control Groups					
Random	23	0.45*	Yes	0.32 to 0.58	n.s.
Nonequivalent Groups	39	0.43*	No	0.37 to 0.50	
Existence of Pretreatment Group Differences					
Present	5	0.13n.s.	Yes	-0.08 to 0.35	n.s.
Absent	41	0.47*	No	0.39 to 0.54	
Present but Adjusted	6	0.48*	Yes	0.36 to 0.60	
Not Given	10	0.40*	Yes	0.24 to 0.56	
Sample Size					
20 to 31	14	0.48*	No	0.26 to 0.70	n.s.
32 to 52	16	0.31*	Yes	0.15 to 0.47	
53 to 79	16	0.36*	No	0.23 to 0.49	
80 to 320	16	0.49*	No	0.41 to 0.57	

[a] Effect sizes were tested statistically.* Indicates that an effect size was significantly greater than zero at $p < .05$; n.s. indicates it was not significantly different from zero.

[b] Effect sizes were tested statistically for homogeneity using the *Q* statistic. "Yes" indicates that set was homogeneous and "No" that it was not at $p < .05$.

[c] Pairs of effect sizes for levels of moderators were tested statistically to determine whether they differed from each other at $p < .05$; n.s. means that the differences were not statistically significant.

[d] The same six comparisons contributed effect sizes to both outcomes.

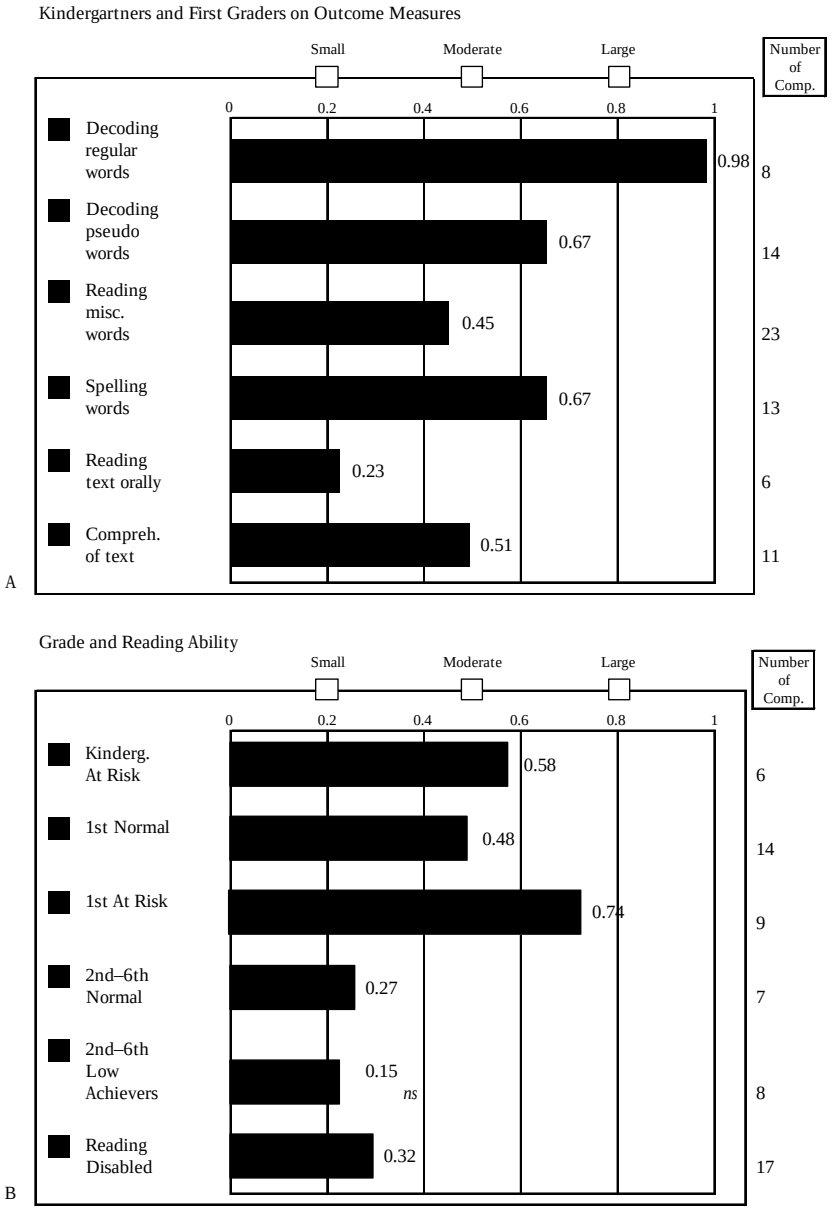
[e] NA = normally achieving readers; RD = students with reading disability; LA = low achieving readers.

[f] This effect size was adjusted to reduce the impact of one atypically large outlier.

[g] Letters in parentheses refer to the type of phonics program: S = synthetic, LU = Larger units.

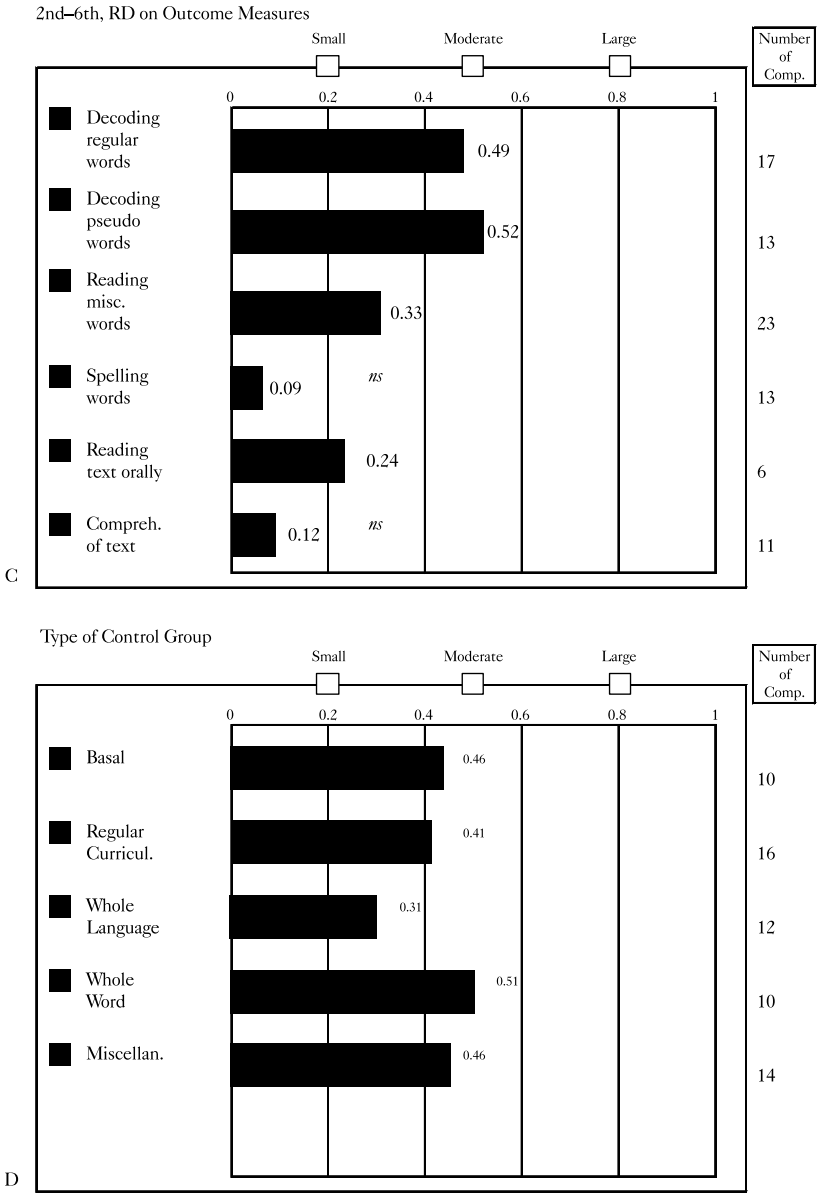
Figure 1

Depiction of mean effect sizes on the overall reading outcome and on specific reading and spelling outcomes as a function of grade, reading ability, and type of control group. All effect sizes were statistically greater than zero except those marked ns *(not significant).*



continued on next page

Figure 1, continued



Note. Compreh. = Comprehension; Comp. = Comparisons.

Effect sizes were calculated on the general reading outcome measured at three possible test points: (a) at the end of instruction or at the end of one year if instruction lasted longer; (b) at the end of instruction; (c) at follow-up points after a delay ranging from 4 months to 1 year. Table 1 reveals that the mean effect size of phonics instruction on reading was statistically greater than zero and moderate in size. When effects of programs were tested at their conclusion, $d = 0.41$. When outcomes were measured at the end of the program or at the end of the first school year for programs lasting longer, $d = 0.44$. Figure 2 shows a stem and leaf plot of the entire pool of effect sizes. It is apparent that most instruction lasted no more than a year and that effect sizes varied substantially. Most (89%) of the 1-year effect sizes were positive. About one-third (31%) were smaller than 0.20. For instruction lasting longer than a year, 80% of the effect sizes were larger than 0.20. On follow-up tests, 86% of the effect sizes ranged from 0.28 to 0.86, indicating that effects lasted beyond the end of instruction. These findings support the conclusion that systematic phonics helps children learn to read more effectively than programs with little or no phonics instruction. Consistent with Figure 2, the pool of end-of-training/first year effect sizes was not found to be homogeneous, opening the possibility that moderators might explain the variation.

Six comparisons assessed both immediate and long-term effects of phonics instruction, with delays ranging from 4 months to 1 year after instruction ended. As shown in Table 1, the effect size remained statistically greater than zero but declined from $d = 0.51$ immediately after instruction to $d = 0.27$ at follow-up. Both sets of effect sizes were homogeneous. This shows that the impact of phonics instruction lasted well beyond the end of training.

Treatment–control comparisons were coded for various characteristics to determine whether phonics instruction was effective under different circumstances. Unless stated otherwise, the primary outcome used to analyze moderator effects was the pooled measure of reading taken at the end of phonics instruction or at the end of the first school year when the program continued beyond this. It is important to recognize that this type of analysis of moderator variable effects has limitations. When differential effects are detected under specific circumstances, one cannot be sure that the circumstances in focus were responsible rather than some other factor confounded with that circumstance. For example, if an effect size associated with tutoring appears to be smaller than an effect size for small groups, this might arise not because tutoring is less effective but rather because the hardest-to-teach students were disproportionally represented in the tutoring studies. Thus, causal inferences involving moderators remain tentative and are merely suggestive of further studies to be done.

Inspection of the column of effect sizes associated with moderator variables in Table 1 reveals that the vast majority were significantly greater than zero (those marked with an asterisk). This suggests that systematic phonics instruction was effective across a variety of conditions and characteristics. However, some findings are based on relatively few comparisons rendering conclusions more tentative.

Specific Reading/Spelling Outcomes. The aim of phonics instruction is to help children acquire alphabetic knowledge and use it to read and spell words. From Table 1, it is apparent that effect sizes for all six types of outcomes were statistically greater than zero, indicating that phonics instruction benefited not only word reading and spelling but also text processing. Effects were strongest on measures of decoding regularly spelled words ($d = 0.67$) and pseudowords ($d = 0.60$). These effects were statistically larger than effects observed on the other four measures that did not differ statistically from each other. This indicates that phonics instruction was especially effective in teaching children to decode novel words, one of the main goals of phonics.

Table 2

Unweighted Statistics as a Function of Moderator Variables With Outcomes Measured Immediately After Training or at the End of One School Year When Training Lasted Longer Unless Otherwise Labeled

Moderator Variable and Level	Number Cases	Mean *g*	Median *g*	*SD*	Minimum/ Maximum *g*
Time of Posttest					
End of Training	65	0.48	0.38	0.66	-0.48 / 3.76
End of Training or First Year	62	0.50	0.38	0.68	-0.48 / 3.76
End of Training[a]	6	0.50	0.53	0.27	0.04 / 0.78
Follow-up[a]	6	0.32	0.33	0.45	-0.48 / 0.88
Outcome Measures					
Decoding regular words	30	0.70	0.60	0.69	-0.16 / 3.33
Decoding pseudowords	39	0.67	0.59	0.42	-0.11 / 1.75
Reading misc. words	59	0.42	0.38	0.61	-1.2 / 2.97
Spelling words	37	0.38	0.26	0.60	-0.58 / 2.76
Reading text orally	16	0.85	0.24	2.23	-0.41 / 8.90
Comprehending text	35	0.31	0.29	0.49	-0.82 / 1.88
Characteristics of Participants					
Grade Levels					
Kindergarten & 1st	30	0.70	0.50	0.83	-0.35 / 3.76
2nd–6th (NA, RD, LA[b])	32	0.30	0.25	0.41	-0.48 / 1.47
Younger Grades					
Kindergarten	7	0.54	0.51	0.17	0.33 / 0.74
1st Grade	23	0.75	0.49	0.95	-0.35 / 3.76
Grade and Reading Ability					
Kindergarten At Risk	6	0.54	0.56	0.19	0.33 / 0.74
1st Normal Achieving	14	0.61	0.39	0.85	-0.35 / 2.83
1st At Risk	9	0.98	0.64	1.09 >	0.08 / 3.76
2nd–6th Normal Achieving	7	0.20	0.38	0.36	-0.48 / 0.61
2nd–6th Low Achieving	8	0.16	0.13	0.26	-0.21 / 0.64
Reading Disabled	17	0.41	0.27	0.46	-0.26 / 1.47
Outcome Measures					
Kindergartners and First Graders					
Decoding regular words	8	0.96	0.78	1.05	-0.06 / 3.33
Decoding pseudowords	14	0.70	0.62	0.38	0.08 / 1.51
Reading miscellaneous words	23	0.57	0.44	0.87	-1.2 / 2.97
Spelling words	13	0.79	0.64	0.80	-0.58 / 2.76
Reading text orally	6	1.91	0.14	3.54	0.00 / 8.90
Comprehending text	11	0.59	0.63	0.61	-0.46 / 1.88
2nd–6th (NA, RD, LA[b])					
Decoding regular words	17	0.61	0.57	0.57	-0.16 / 2.14
Decoding pseudowords	13	0.66	0.53	0.52	-0.11 / 1.75
Reading miscellaneous words	23	0.29	0.32	0.40	-0.50 / 1.07
Spelling words	13	0.04	0.05	0.25	-0.26 / 0.46
Reading text orally	6	0.16	0.24	0.27	-0.29 / 0.45
Comprehending text	11	0.14	0.12	0.37	-0.62 / 0.61
Socioeconomic Status					
Low SES	6	0.65	0.68	0.37	0.08 / 1.23
Middle SES	10	0.47	0.22	0.86	-0.11 / 2.83
Varied	14	0.41	0.37	0.26	0.03 / 0.92
Not Given	32	0.51	0.38	0.79	-0.48 / 3.76

continued on next page

Table 2, continued

Moderator Variable and Level	Number Cases	Mean *g*	Median *g*	*SD*	Minimum/ Maximum *g*
Characteristics of Instruction					
Type of Phonics Program					
Synthetic	39	0.51	0.43	0.59	-0.48 / 2.83
Larger Phon. Units[c]	11	0.70	0.37	1.11	-0.26 / 3.76
Miscellaneous	10	0.22	0.21	0.30	-0.35 / 0.63
Specific Phonics Programs[d]					
NRS-Beck LRDC (S)	4	0.48	0.45	0.16	0.33 / 0.71
Direct Instruction (S)	4	0.85	0.65	0.99	0.01 / 2.08
Lovett Direct Instruct (S)	4	0.51	0.25	0.64	0.09 / 1.47
Lovett Analogy (LU)	4	0.55	0.51	0.70	-0.26 / 1.46
Lippincott (S)	3	0.61	0.51	0.19	0.49 / 0.84
Orton-Gillingham (S)	10	0.22	0.17	0.27	-0.21 / 0.64
Sing Spell Read Write (S)	3	0.38	0.38	0.13	0.26 / 0.51
Instructional Delivery Unit					
Tutor[c]	8	1.09	0.54	1.22	0.33 / 3.76
Small Group	27	0.44	0.44	0.42	-0.26 / 1.47
Class	27	0.37	0.26	0.60	-0.48 / 2.83
Type of Control Group					
Basal	10	0.57	0.48	0.28	0.26 / 1.23
Regular Curriculum	16	0.52	0.36	0.90	-0.21 / 3.76
Whole Language	12	0.26	0.19	0.44	-0.48 / 0.92
Whole Word	10	0.53	0.32	0.86	-0.11 / 2.83
Miscellaneous	14	0.59	0.45	0.64	-0.26 / 2.08
Characteristics of the Design of Studies					
Assignment of Participants to Treatment and Control Groups					
Random	23	0.53	0.40	0.54	-0.26 / 2.08
Nonequivalent Groups	39	0.47	0.38	0.75	-0.48 / 3.76
Existence of Pretreatment Group Differences					
Present	5	0.11	0.04	0.13	0.00 / 0.27
Absent	41	0.59	0.40	0.79	-0.48 / 3.76
Present but Adjusted	6	0.46	0.45	0.25	0.08 / 0.74
Not Given	10	0.33	0.31	0.34	-0.21 / 0.84
Sample Size					
20 to 31	14	0.58	0.49	0.73	-0.35 / 2.08
32 to 52	16	0.31	0.25	0.25	-0.21 / 0.78
53 to 79	16	0.48	0.33	0.93	-0.48 / 3.76
80 to 320	16	0.62	0.48	0.63	-0.11 / 2.83

[a] The same six comparisons contributed effect sizes to both outcomes.

[b] NA = normally achieving readers; RD = students with reading disability; LA = low achieving readers.

[c] This effect size was adjusted to reduce the impact of one atypically large outlier.

[d] Letters in parentheses refer to the type of phonics program: S = synthetic, LU = larger units.

Figure 2

Stem-and-leaf plot showing the distribution of mean effect sizes of systematic phonics instruction on reading measured at the end of instruction and following a delay.

Stem	Leaf End of Instruction (1 yr)[a]	Leaf End of Instruction (>1 year)[b]	Leaf Followup[c]
3.7	1		
2.2	7		
2.1			
2.0			
1.9	9		
1.8			
1.7			
1.6			
1.5			
1.4	12		
1.3			
1.2			
1.1	9		
1.0			
0.9	1		
0.8	4		
0.7	0236	5	6
0.6	001233	47	
0.5	00133	24	6
0.4	345789		
0.3	23367889	6	238
0.2	014457	48	8
0.1	23469	7	
+0.0	01344479	0	
-0.0	7		
-0.1	1		
-0.2	05		
-0.3	3		
-0.4	7		7

[a] End of instruction or end of Year 1 when instruction lasted longer.

[b] End of instruction which lasted between 2 and 4 years.

[c] Followup tests were administered 4 to 12 months after instruction ended.

Characteristics of Students. The students who received phonics instruction varied in age/grade in school. Kindergartners and first graders, particularly those at risk, typically began phonics instruction as nonreaders or novice readers with much to learn, whereas children in second through sixth grades had already been exposed to reading instruction and had made at least some progress when phonics instruction was introduced. Most of the comparisons with older students (78%) involved readers with RD or low achieving readers. Of interest were whether phonics instruction made a contribution across groups and whether its impact was larger when introduced early.

From Table 1, it is apparent that phonics instruction facilitated reading acquisition in both younger and older readers. Effect sizes were statistically greater than zero. However, as predicted, the impact was statistically larger among beginners (d = 0.55) than among older children (d = 0.27).

Effects were moderate and very similar for kindergartners (d = 0.56) and first graders (d = 0.54). Many more comparisons occurred in first grade than in kindergarten, making the first grade findings more reliable. These findings support Chall's (1996b) claim that phonics instruction should exert its greatest impact early.

In most of the studies, phonics instruction lasted 1 school year or less. However, there were four treatment–control comparisons where phonics instruction began in kindergarten or first grade and continued for 2 or 3 years (Blachman et al., 1999; Brown & Felton, 1990; two comparisons from Torgesen et al., 1999). These studies were all conducted with at-risk readers. Mean effect sizes at the end of each grade level were moderate and their strength was maintained across the grades: kindergarten d = 0.46; first grade d = 0.54; second grade d = 0.43. This offers further support for the value of starting phonics early and continuing to teach it for 2 to 3 years.

To clarify whether phonics instruction was effective for various types of readers, treatment–control group comparisons were grouped by grade and reading ability. As shown in Table 1 and Figure 1A, statistically significant effect sizes were evident for all but one group. Mean effect sizes were moderate to large, ranging from $d = 0.48$ to $d = 0.74$, among at-risk and normally achieving readers in kindergarten and first grades. Effect sizes were significant but smaller for second through sixth grade normally achieving readers ($d = 0.27$) and students with RD ($d = 0.32$). These findings indicate that phonics instruction improves reading ability more than nonphonics instruction not only among beginning readers but also among normally progressing readers above first grade and older readers with RD. In contrast, phonics instruction did not enhance reading among low achieving older readers ($d = 0.15$).

Because effects of phonics instruction on overall reading varied among these reader groups, we wondered whether the groups might also vary on the specific outcome measures. Results in Table 1 and Figure 1B show that, among beginners, phonics instruction produced statistically higher reading performance on all six measures. The effect size on decoding regularly spelled words was large ($d = 0.98$) and statistically greater than nondecoding measures. A moderately large effect was also produced on measures of decoding pseudowords ($d = 0.67$). Large effects are not surprising because a central goal of phonics programs is teaching students to decode. The effect size for reading miscellaneous words ($d = 0.45$) was somewhat reduced, perhaps because these tests often included irregularly spelled words not amenable to decoding. The impact of phonics instruction on reading comprehension ($d = 0.51$) was moderate and similar to that for reading miscellaneous words, perhaps reflecting the fact that beginners' text reading is heavily dependent upon their familiarity with the words. The stem and leaf display in Figure 3 shows the distribution of effect sizes on the comprehension outcome for beginners. It reveals that all but one *d* was positive, hence bolstering the conclusion that systematic phonics instruction helps beginning readers comprehend text.

We expected phonics instruction to facilitate spelling in beginners, particularly because spelling was often measured by the ability to apply grapheme–phoneme correspondences to write sound spellings of words. Several studies employed scoring systems that gave credit for phonetically plausible though incorrect

Figure 3

Stem-and-leaf plot showing the distribution of mean effect sizes of systematic phonics instruction on reading comprehension measured at end of instruction as a function of grade and reading level.

Stem	Leaf Kindergarten/ First Grade	Leaf 2nd–6th Grades Normal Ach.	RD	LA
1.8	0			
1.0	8			
0.9				
0.8	3			
0.7	36			
0.6	0		02[a]	
0.5			7	
0.4	9		2	
0.3	266		6	
0.2	89		7	
0.1	0178		2	
+0.0			79	
-0.0				
-0.1				9
-0.2				4
-0.3				
-0.4	3			
-0.5				
-0.6		1		

[a] This effect size was measured at the end of instruction lasting 2 years (Oakland et al., 1998).

spellings, for example, spelling *feet* as FET or *car* as KR. This may have contributed to the sizeable effect observed on the spelling outcome (d = 0.67) in beginners.

Among the older readers in second through sixth grades, a somewhat different picture emerged. Although phonics instruction produced a small but statistically significant effect on the overall measure of reading (d = 0.27), effects on specific outcomes were not uniformly small. As evident in Table 1 and Figure 1C, superior performance decoding regularly spelled words (d = 0.49) and pseudowords (d = 0.52) was moderate and statistically greater than zero. Effects of phonics instruction were somewhat smaller but statistically significant in reading miscellaneous words (d = 0.33) and reading text orally (d = 0.24). In contrast, phonics instruction was not effective for teaching spelling (d = 0.09) or teaching reading comprehension (d = 0.12).

Because readers in second through sixth grades classified as low achieving (LA) revealed no overall effects of phonics instruction in contrast to normally progressing readers and students with RD, we examined whether the above pattern of specific outcomes held when effect sizes for the three reader groups (NA, RD, and LA) were examined separately. In the analysis of RD comparisons, effect sizes proved almost identical to the overall pattern in Figure 1C except on one outcome. The effect size on reading comprehension was statistically greater than zero (d = 0.27, based on eight comparisons that were homogeneous). This indicates that systematic phonics instruction did improve the ability of students with RD to comprehend text. There were too few comparisons in the other groups to conduct a similar analysis.

The stem and leaf plot of effect sizes on the comprehension measure presented in Figure 3 reveals that all the effect sizes were positive for readers with RD, whereas effect sizes were negative for normally progressing and LA readers. These findings reveal that the nonsignificant effect on the comprehension outcome among second- through sixth-grade students arose primarily from the students without RD.

One additional characteristic of children was examined, their socioeconomic (SES) status. Table 1 shows that effect sizes were statistically greater than zero. Phonics instruction exerted its strongest impact on low SES children (d = 0.66). However, effect sizes did not differ statistically among the four SES categories (i.e., low, middle, varied, not given). These findings indicate that phonics instruction contributes to higher performance in reading in both low and middle class students.

Characteristics of Phonics Instruction. The treatment–control group comparisons were categorized by the type of systematic phonics instruction taught. Synthetic phonics programs taught students to transform letters into sounds (phonemes) and to blend the sounds to form recognizable words. This was by far the most common type of program, used in 43 comparisons with 39 comparisons measuring outcomes at the end of instruction. Some of these programs were developed by researchers whereas others were published programs, for example, Jolly Phonics (Lloyd, 1993), Lindamood ADD program (Lindamood & Lindamood, 1984), Lippincott Basic Reading (1981), Open Court Reading (1995), Orton Gillingham (Gillingham & Stillman, 1979), Direct Instruction/*Reading Mastery*/*DISTAR* (Engelmann, 1980), and Sing Spell Read & Write (Dickson, 1972).

In 11 comparisons children were taught to analyze and blend larger subunits of words such as onsets, rimes, phonograms, or spelling patterns along with phonemes. Embedded code programs taught blending in the context of words and text. In some programs, rimes provided the basis for teaching students to read new words by analogy to known words. Words in texts were built from linguistic patterns. Writing complemented reading in most programs. The database included Hiebert's embedded code program (Hiebert, Colt, Catto, & Gary, 1992), Reading Recovery Programs (Clay, 1993) that were modified to

include systematic phonics, and a program derived from the Benchmark Word Identification program (Gaskins et al., (1988). One of the studies, by Tunmer and Hoover (1993), exhibited an atypical effect size, d = 3.71. To limit its influence on the mean effect size for the larger-unit phonics category, its effect size was reduced to equal the next largest effect size in the set, d = 1.41.

The Miscellaneous category consisted of phonics programs that could not be categorized. This set included a spelling program, traditional phonics basal programs, and some researcher-devised instruction that focused on word analysis procedures.

Two comparisons combined a synthetic program with a larger-unit word analogy program (Lovett et al., 2000). Their mean effect size was d = 0.42. They were not included in the analysis of phonics types.

As evident in Table 1, the three categories of systematic phonics programs produced effect sizes that were statistically greater than zero, showing that they were more effective than nonsystematic phonics programs in helping children learn to read. The synthetic phonics programs exerted a moderate impact on reading (d = 0.45). A slightly smaller impact was produced by the larger-unit programs (d = 0.34) and the miscellaneous programs (d = 0.27). However, the three effect sizes did not differ statistically ($p > .05$).

Seven phonics programs were evaluated in three or more treatment–control comparisons (see program descriptions in Appendix A). Results in Table 1 reveal that all produced effect sizes statistically greater than zero and none differed statistically from the others. Effect sizes ranged from a high of d = 0.68 for the Lippincott program to a low of d = 0.23 for the Orton-Gillingham program. One reason for the lower value may be that Orton-Gillingham comparisons were tested with older students, many of whom were poor readers. These findings suggest that all the specific phonics programs helped children learn to read better than control group programs. However, drawing the conclusion that these programs are all equally effective is dubious because there were very few comparisons assessing each program, so power was low.

Another property of systematic phonics instruction expected to influence performance in reading was the delivery unit. Tutoring one-on-one was expected to be most effective, particularly for low achieving readers and students with RD, because it was tailored to individual students. Small group instruction was also expected to be effective, because attention to individual students is still possible, and in addition, the social setting was expected to enhance motivation to perform and opportunities for observational learning.

In categorizing studies, we found it easy to determine that tutoring or small groups were used, although it was not always clear that this was the only form of instruction delivered. In contrast, whole class categorization was more difficult. In many reports, descriptions stated that teachers taught the phonics program to their students, but the delivery unit was not explicitly stated. In these cases, we inferred it to be the class.

Before the meta-analysis was conducted, an adjustment was made to one atypically large effect size in the tutoring set (e.g., Tunmer & Hoover, 1993, effect size of d = 3.71 reduced to the next largest effect size in the set, d = 1.99).

Results revealed that all three forms of instruction produced positive effects that were statistically greater than zero, indicating that tutoring, small groups and classes were all effective ways to deliver phonics instruction to students (see Table 1). Although tutoring produced a slightly larger effect size (d = 0.57), it did not differ statistically from the other effect sizes (d = 0.39 and 0.43). This evidence falls short in showing that tutoring was more effective for teaching phonics. The fact that most of the control groups against which tutoring was compared (i.e., 62%) did not receive tutoring should have given tutoring an extra advantage,

but it did not. Also, it was not the case that tutoring was employed with hard-to-teach students. Inspection of the types of students who received tutoring revealed that six were kindergartners or first graders whereas only two were older poor readers.

Inspection of effects for individual studies (see Appendix B) revealed that some whole class programs produced effect sizes as large or larger than those produced by small groups or tutoring. The fact that classroom instruction can be as effective as tutoring is important to note given the expense and impracticality of delivering instruction individually.

Characteristics of Control Group Instruction. Control groups varied in whether unsystematic or incidental phonics was taught or instruction lacked any attention to phonics. Control groups were categorized as one of five types based on labels or descriptions provided by authors: basal, regular curriculum, whole language, whole word, miscellaneous.

Basal programs typically were those already in use at schools. "Regular curriculum" covered cases where controls received the traditional curriculum or the regular class curriculum in use at schools with no further specification of its contents other than asserting it did not teach phonics systematically. This category included cases where performance in the same grade at the same school during previous years was used as a baseline without additional description of the program taught. This also included regular intervention programs provided by schools to treat reading problems in studies involving struggling readers.

Programs were classified as whole language when authors used this label to characterize instruction. These included programs using Big Books (Holdaway, 1979) and language experience. For example, Stuart (1999) provided oversized books for kindergarten teachers to read and identical smaller books for the students to practice repeated reading. Teachers were encouraged to spend time on word level work by discussing words and letters in the texts but no system was followed. In the Foorman, Francis, Fletcher, and Schatschneider (1998) study, the district's whole language program was taught to the control group. Predictable books and writing activities were included, with teachers using shared and guided reading activities to draw children's attention to specific words, letters, sounds, and patterns. The act of making meaning from print was the focus of instruction, and learning the alphabetic code occurred incidentally. In two studies (Freppon, 1991; Klesius, Griffith, & Zielonka, 1991), the purpose was to examine the effectiveness of whole language programs, not phonics which was taught using a "skill and drill" basal program that was not well described. Whole language programs were taught to control groups primarily in studies involving children in first grade (67% of the comparisons).

Studies in which the control groups received whole word or sight word instruction without much attention to letter–sound relations were classed as whole word programs. For example, in one study (Brown & Felton, 1990), control children were taught to identify words by using context cues and then verifying that the words have the "right sounds." Although phonics elements were covered by the end of second grade, children were not shown how to use them to decode words. The primary emphasis was on acquiring a sight vocabulary by memorizing whole words. In another study (Fulwiler & Groff, 1980), the control group received a "less intensive phonics" approach in which 50–75 sight words were taught before incidental phonics instruction began.

Control groups receiving instruction that did not fit into our scheme were placed in the Miscellaneous category. These included programs teaching traditional spelling, academic study skills, and tutoring in academic subjects. In one study examining the effectiveness of parents teaching systematic phonics to their own children, parents in the control group spent time listening to their children read books (Leach & Siddall, 1990). Many of these programs appeared to teach no phonics.

Results in Table1 and Figure 1D reveal that effect sizes favoring the phonics treatment were statistically greater than zero for all types of control groups. Mean effect sizes ranged from $d = 0.31$ for whole language controls to $d = 0.51$ for whole word controls. None of the effect sizes differed statistically from the others. These findings suggest that phonics instruction produced superior performance in reading regardless of the type of control group that was used.

Characteristics of Research Design. Studies in the database varied in methodological rigor. Three features were coded and analyzed to determine whether more rigorous designs yielded larger or smaller effect sizes: use of random assignment; existence of pretreatment group differences; and sample size. The variation among studies on all of these dimensions showed that some studies were stronger methodologically than others.

Most studies in the database provided information regarding whether students were assigned randomly to treatment and control groups. If this was not mentioned, then the study was categorized as using nonequivalent groups. Table 1 shows that only 37% of the studies used random assignment. However, both types of assignment yielded similar effect sizes, and both effects were statistically greater than zero: $d = 0.45$ (random) and $d = 0.43$ (nonequivalent). These findings show that the positive effects of phonics instruction on reading did not arise primarily from studies with weaker nonequivalent group designs.

Studies were coded for the presence or absence of pretest differences between treatment and control groups. From Table 1, it is apparent that only five comparisons (8%) showed this methodological weakness, with pretest differences favoring the phonics treatment group over the control group in all cases.[2] Only 10 comparisons (16%) failed to report pretest information. The majority of the comparisons showed rigor in this respect (76%). It is interesting that, as shown in Table 1, the more rigorous studies revealed statistically significant effects, $d = 0.47$. In contrast, the mean effect size of studies with pretreatment differences was not statistically greater than zero, $d = 0.13$. These findings reveal that the positive benefits of phonics instruction did not arise from poorly designed experiments in which the better students received the phonics treatment.

Another characteristic involving design rigor was sample size. Studies with larger sample sizes are more highly regarded because their findings are more reliable. Typically in meta-analyses, studies with larger samples show smaller effect sizes (Johnson & Eagly, 2000). Sample sizes in our treatment–control comparisons showed great variation, ranging from 20 to 320. We grouped the comparisons into four blocks to compute mean effect sizes. From Table 1, it is apparent that effect sizes were all statistically greater than zero, were similar across the blocks, and did not differ statistically from each other. Surprisingly, the largest effect size, $d = 0.49$, emerged in studies using the largest samples, which is contrary to the typical finding. These results show that the positive effects of phonics instruction were not limited to studies that produced effects with relatively few students.

Illustrative Experiments

Eight studies in the database serve to illustrate the experiments contributing effect sizes. These studies were selected to portray different types of phonics instruction taught to kindergartners, first and second graders, and older students with RD in well designed studies that delivered instruction to individuals, small groups, and classrooms.

Systematic Phonics Instruction in Kindergarten. Phonics instruction appropriate for children in kindergarten, particularly English Language Learners (ELL), was studied by Stuart (1999), who compared the Jolly Phonics program (Lloyd, 1993) to a whole language approach using big books (Holdaway, 1979). Three teachers taught each program 1 hr per day for

12 weeks to at-risk children during the latter half of kindergarten. The sample included 86% ELL students whose first language was not English. This was the only study in the database that examined whether phonics instruction is effective with ELL students.

The Jolly Phonics program is systematic and prescribed in its teaching of letters. The program uses meaningful stories, pictures, and actions to reinforce recognition and recall of letter–sound relationships and precise articulation of phonemes. The program includes five key elements: (a) learning the letter sounds, (b) learning letter formation, (c) blending sounds for reading, (d) identifying the sounds in words for writing, and (e) learning tricky words that are high frequency and irregularly spelled. Unlike many older phonics approaches, Jolly Phonics incorporates playful, creative, flexible teaching to lead children from early literacy skills to authentic reading and writing.

Big Book instruction includes work with letters. Teachers draw children's attention to written words in the books and they talk about letters in words. Also teachers employ various "imaginative and fun activities" to help children learn letters and their sounds. However, the instruction is not systematic, the sequence of teaching letters is not prescribed, and no special system for remembering letter–sound relations is taught.

At the end of training, test results showed that Jolly Phonics children were able to read significantly more words and pseudowords and to write more words than Big Book children. The overall effect size was $d = 0.73$. In a follow-up test 1 year later, the phonics group outperformed the control group in reading and spelling words but not in reading comprehension. This may have occurred because most of the students were ELL. For their comprehension to be improved, perhaps more extensive instruction to enhance competency in English syntax and semantics is required.

One interesting feature of the Jolly Phonics program is that children are taught hand gestures to help remember the letter–sound associations. For example, they make their fingers crawl up their arm to portray an ant while they chant the initial sound /ae/ of "ant" associated with the letter A. Another kindergarten program also uses mnemonics to teach letter–sound relations. In Letterland (Wendon, 1992), all the letters are animated characters that assume the shapes of the letters and have names prompting the relevant sound, for example, Sammy Snake, Hairy Hat Man, Fireman Fred, Annie Apple.

The task of learning the arbitrary, meaningless shapes, names and sounds of all the alphabet letters is difficult and time-consuming, particularly for children who come to school knowing few letters. Techniques to speed up the letter learning process are valuable in helping kindergartners prepare for formal reading instruction. Letter knowledge is one of the two best kindergarten predictors of how well children will learn to read, the other being phonemic awareness (Share, Jorm, Maclean, & Matthews, 1984). The value of mnemonics for teaching letter–sound relations was verified in a study showing that Letterland-type mnemonics helped children learn letter–sound relations better than mnemonics which did not link letter shapes to sounds (Ehri, Deffner, & Wilce, 1984).

A Longitudinal Classroom Study of Systematic Phonics. A lengthy, comprehensive program lasting over 2 years was studied by Blachman et al. (1999). Classroom teachers taught the program to low SES, inner-city children. Instruction began in kindergarten with a focus on phonemic awareness and lasted 11 weeks. In first grade, explicit, systematic instruction in the alphabetic code was taught. During second grade, children who had not met the program's goals received additional instruction while the rest received regular classroom instruction. Control children participated in the school's regular basal reading program that included a phonics workbook that children used independently.

During phonemic awareness instruction, children were taught to perform a "say it and move it" procedure in which they moved a disk down a page as they pronounced each phoneme in two- and three-phoneme words. Then a limited set of eight letter–sound relations was taught and children moved the letters rather than the disks. It is noteworthy that when children began this program they started from zero in their alphabetic learning. They knew on average only two letter–sounds and could not write their names. However, by the end of kindergarten, children knew on average 19 letter names and 13 letter sounds.

Despite having received the same instruction, children in the treatment group varied in their letter knowledge and phonemic awareness at the beginning of first grade. To address the variation, they were grouped by instructional levels. The core of the reading program involved daily, 30-minute lessons consisting of five steps that emphasized the alphabetic code:

1. teaching new sound–symbol correspondences with vowels highlighted in red;
2. teaching phoneme analysis and blending;
3. reading regularly spelled, irregularly spelled, and high-frequency words on flash cards to develop automaticity;
4. reading text containing phonetically controlled words;
5. writing four to six words and a sentence to dictation.

By the end of the program, children had been introduced to all six syllable types: closed (*fat*), final E (*cake*), open (*me*), vowel team (*pain*), vowel +r (*burn*), and consonant *le* (*table*). Vocabulary development and work on reading comprehension were incorporated as well, with more time spent reading text as the year progressed and children's reading vocabulary grew.

Inservice workshops on program implementation were held once a month. Teachers learned how children acquire literacy skills and the role of phonological processes in learning to read. They learned how to provide explicit instruction in the alphabetic code. The issue of pacing was stressed. Developing students' phonemic awareness, letter–sound knowledge, and word recognition skills were identified as being more important than "covering the material."

Children's progress in reading and writing was assessed at the end of each grade for both treatment and control groups. Results showed moderate to large effect sizes favoring the phonics group: $d = 0.72$ (kindergarten), $d = 0.64$ (first grade), and $d = 0.36$ (second grade). These findings illustrate how explicit systematic instruction in phonics improved low SES children's ability to read when instruction was provided by classroom teachers and when instruction continued from kindergarten through second grade.

An Intensive Three-Year Tutoring Program. Torgesen et al. (1999) compared two types of phonics instruction delivered by tutors to children at risk for future reading difficulties. The program lasted from kindergarten through second grade. One program provided explicit and intensive instruction in phonemic awareness and decoding called PASP (phonological awareness plus synthetic phonics) while the other program provided systematic but less explicit instruction in phonemic decoding in the context of more instruction and practice in text comprehension, called EP (embedded phonics). Kindergarten children with poor phonemic awareness and letter knowledge received 88 hr of tutoring over 2.5 years, in sessions lasting 20 min and scheduled four times per week. Instruction was individually paced and occurred in addition to classroom reading instruction.

Two control groups were used, one that received tutoring in the methods and materials being used to teach reading in the regular classroom, and one receiving only regular classroom instruction. Some phonics oriented activities were included. There were 180 children

from 13 schools, with children randomly assigned to one of the four conditions.

The PASP children received the Auditory Discrimination in Depth program (Lindamood & Lindamood, 1984). This program taught children phonemic awareness in a unique way. They were led to discover and label the articulatory gestures associated with each phoneme by analyzing their own mouth movements as they produced speech. For example, children learned that the word *beat* consists of a lip popper, a smile sound, and a tongue tapper. Children learned to track the sounds in words with mouth pictures as well as colored blocks and letters. Much time was spent building children's phonemic awareness and their decoding skills but some attention was given to high frequency word recognition, text reading, and comprehension.

The Embedded Phonics (EP) program began by teaching children to recognize whole words. Instruction in letter–sound occurred in the context of learning to read words from memory. Also children wrote sentences and read what they wrote. Phonemic awareness was taught by having children segment the sounds in words before writing them. When children had sufficient reading vocabulary, they began reading short stories to build their reading vocabulary further. The emphasis was on acquiring word level reading skills, including sight words and phonemic decoding skills. Also, attention was given to constructing the meanings of stories that were read.

Researchers videotaped 25% of the PASP and EP tutorial sessions to verify that phonics instruction differed in the two programs. The percentages of time spent on the following activities were:

- Phonemic awareness, letter–sounds, phonemic reading/writing of words: 74% (PASP) vs. 26% (EP);
- Sight word instruction: 6% (PASP) vs. 17% (EP);
- Reading/writing connected text: 20% (PASP) vs. 57% (EP).

Statistical tests comparing performance on outcomes across the grades revealed that the PASP group read significantly more real words and nonwords and spelled more words correctly than one or both of the control groups. In contrast, the EP group did not outperform the control groups on any of these measures. None of the groups differed statistically in reading comprehension. When each phonics group was compared to the classroom control group on the overall measure of reading, the following effect sizes were observed:

- PASP: d = 0.33 (kindergarten), 0.75 (first grade), 0.67 (second grade);
- EP: d = 0.32 (kindergarten), 0.28 (first grade), 0.17 (second grade).

Although intensive synthetic phonics instruction helped at-risk children learn to read words better than embedded phonics instruction that produced little advantage over control instruction, no benefit to comprehension was observed. This was surprising given that phonics training lasted for 2.5 years, the EP group received substantial text-based instruction, and children began the programs as nonreaders with much room to grow. Admittedly, reading comprehension depends upon other processes besides word reading, but one would expect to see a benefit, particularly in the primary grades when text reading is heavily influenced by word recognition skills. One possible explanation is that the tests of comprehension were standardized and so may not have been sufficiently sensitive to detect small within-grade differences. Swanson (1999) found that standardized comprehension tests yielded smaller effect sizes than experimenter-devised comprehension tests.

Enriching Reading Recovery With Phonics. The Reading Recovery (RR) program developed by Clay (1993) provides tutoring to children who have fallen behind in reading after a year of instruction. The 30-min RR lesson includes several activities: rereading two familiar books,

reading the previous day's new book, practicing letter identification, writing a story by analyzing sounds in words, re-assembling the words of a cut-up story, and reading a new book.

Greaney, Tunmer, and Chapman (1997) modified the RR program by providing explicit instruction in larger-unit rime patterns once children had learned the majority of letters. Unlike most other studies of RR, which have involved younger children, the participants in this study were children from Grades 2 through 5 who were the poorest readers in their class. The phonics lesson consumed 5 min and was substituted for the letter segment in each RR session. Children were taught to read nouns containing common spellings of rimes (e.g., m-eat) and then words with the rime embedded in it (e.g., h-eat-er). They practiced reading and also writing words with these larger rime units which were referred to as "eggs" because the unit was written in an egg-shaped space. During the final book-reading segment of each session, children were encouraged to use the eggs to identify unfamiliar words in the books. Children in the control group followed the same procedures and read the same words, but no attention was drawn to rime units in the words, and the words were mixed up rather than taught in sets having the same rimes.

Treatment and control programs lasted for 12 weeks. Results showed that the children who received rime training outperformed control children on tests of word and pseudowords reading but not on tests of reading comprehension. The overall effect size was $d = 0.37$. These findings indicate that phonics enrichment improves the effectiveness of RR for teaching word reading skills but not reading comprehension skills.

Santa and Hoien (1999) modified the RR format to include more systematic phonics instruction. In their study, at-risk first graders received tutoring that involved story reading, writing, and phonological skills based on a program developed by Morris (1992). The unique part of this phonics program was that it used word study activities to develop phonological awareness and decoding skill. Word study consumed 5–6 min of the 30-min lesson. Children were given cards to sort into categories. They might sort picture cards that shared the same initial sounds, or word cards sharing the same vowel sounds. The typical sort involved three patterns with four words in each pattern. Initially, children worked with phonograms (e.g., -at in *hat, cat, sat, rat*) and then advanced to shared phonemes as the basis for sorting words. Children also were taught to spell by writing letters for the sounds heard in words. Metacognitive strategies were taught including an analogy strategy to read new words.

The control group received small-group, guided-reading instruction. Students practiced reading and rereading books in 30-min lessons but did not receive any word study activities. It is important to note that the control group here was not one that received RR unenriched by phonics. Rather it received a different form of instruction that did not involve tutoring. Results showed that the phonics word study program produced much better performance in reading than the guided reading program, $d = 0.76$. The phonics group significantly outperformed the control group in reading comprehension ($d = 0.73$) as well as word reading ($d = 0.93$). These findings illustrate the effectiveness of larger-unit phonics instruction added to an RR format.

Systematic Phonics Instruction for Students with Reading Disability. Students with severe reading disability have great difficulty acquiring word reading skills, not only decoding but also recognizing words. Systematic phonics programs have been developed to remediate their reading problems. Lovett and her associates (Lovett & Steinbach, 1997; Lovett et al., 2000) explored the effectiveness of two such programs, a synthetic program called PHAB (phonological analysis and blending) and a larger-unit program that teaches children to use subparts of words they know to read new words, referred to as WIST (word identification strategy training).

The PHAB program adopted the Direct Instruction model developed by Engelmann (1980). Children were taught to segment and blend words orally. They were taught letter–sound associations in the context of word recognition and decoding instruction. The program taught a left-to-right decoding strategy to sound out and blend letters into words. Special marks on letters and words provided visual cues to aid in decoding, such as symbols over long vowels and connected letters to identify digraphs. Cumulative, systematic review and many opportunities for overlearning were used. New material was not introduced until the child had fully mastered previously instructed material. Children were taught in small groups.

The WIST program was adapted from the Benchmark Word Identification/Vocabulary Development program created by Gaskins et al. (1988). It taught children how to use four metacognitive strategies to decode words: reading words by analogy, detecting parts of words that are known, varying the pronunciations of vowels to maintain flexibility in decoding attempts, and "peeling off" prefixes and suffixes in words. Children learned a set of 120 key words exemplifying high-frequency spelling patterns, five words per day. They learned to segment the words into subunits so that they could use parts of known words to read other similarly spelled words. They learned letter–sound associations for vowels and affixes. Various types of texts provided children with practice applying the strategies that were taught.

Participants had severe reading problems. They were randomly assigned to the PHAB program, the WIST program, or a nonreading control program teaching academic survival skills. The students ranged in age from 6 to 13 years (second through sixth grades). The three programs took the same amount of time. In one study it was 35 hr, in another study 70 hr.

Lovett's studies contributed four comparisons assessing effects of PHAB and four assessing WIST. The average effect size across the comparisons indicated that both programs produced about the same growth in reading, $d = 0.50$ for PHAB and $d = 0.54$ for WIST. In two of the comparisons, both reading comprehension and word reading were measured. Substantial gains were evident on both measures. These findings illustrate how both synthetic and larger-unit approaches to teaching systematic phonics were found to be effective in helping students with severe reading difficulty improve their reading skills.

Discussion

Influence of Phonics Instruction on Reading

Findings of the meta-analysis support the conclusion that systematic phonics instruction helps children learn to read more effectively than nonsystematic or no phonics instruction. The impact of phonics instruction on reading was significantly greater in the early grades (kindergarten and first grade) when phonics was the method used to start children out than in the later grades (second through sixth grades) after children had made some progress in reading presumably with another method. These results support Chall's (1967) assertion that early instruction in systematic phonics is especially beneficial for learning to read. Although there was some thought that kindergartners might not be ready for phonics instruction (e.g., Chall, 1996b, Stahl & Miller, 1989), findings did not support this. Effect sizes resulting from phonics instruction were similar in kindergarten and first grade.

The impact of phonics instruction was statistically significant but smaller when introduced beyond first grade. Several explanations are possible. One is that other aspects of reading become more important to teach in the later grades. This is suggested in a comparison of effect sizes drawn from the National Reading Panel's report (2000). Whereas phonics instruction produced an effect size of $d = 0.27$ in second through sixth graders, fluency instruction produced an effect size of $d = 0.47$

and some forms of comprehension strategy instruction produced effect sizes above d = 0.80. This suggests that phonics instruction must be coupled with other forms of effective reading instruction in order to achieve maximum impact.

Another explanation is that when phonics instruction is introduced after students have already acquired some reading skill, it may be more difficult to influence how they read because it requires changing students' habits, for example, abandoning the strategy of guessing or cueing on partial letters to read unfamiliar words and adopting the strategy of fully analyzing letters to determine the words identity. Because the first strategy is more easily executed, it may be hard to suppress.

The conclusion that phonics instruction is less effective when introduced beyond first grade may be premature, however. Several mitigating factors may have reduced effect sizes in the studies we examined. The majority of the comparisons in the older group, 78%, involved either low achieving readers or students with RD. Remediating their reading problems may be especially difficult. There were only seven comparisons involving older, normally progressing readers, and four of these came from one study using the Orton-Gillingham method. This method was developed not for upper-elementary-level, normally achieving readers but rather for students with RD. Other types of phonics programs might prove more effective for older readers without any reading problems, for example, phonics programs that improve the decoding of multisyllabic words. Another factor constraining conclusions here is the reliability of these findings, which are based on relatively few comparisons and hence lack statistical power. Thus, determining whether appropriately designed phonics instruction might prove effective for older, normally progressing readers needs further study.

The advantage of systematic over nonsystematic or no phonics instruction was examined in three types of potential or actual poor readers. Moderate to large effect sizes indicated that phonics helped to prevent reading difficulties in beginners at risk for developing reading problems. In fact, effects were significantly greater in first graders at risk for future reading difficulties than in older students who had already become poor readers. This underscores the special importance of teaching phonics early, especially in schools with large numbers of at-risk students who enter school with very little letter knowledge or phonemic awareness.

Systematic phonics instruction produced significant effects among children diagnosed as having a reading disability. Small-to-moderate effect sizes were evident on reading comprehension measures as well as word reading measures. This indicates that systematic phonics is an effective way to remediate reading problems in children whose struggle is specific to reading and does not include more general cognitive difficulties.

In contrast, systematic phonics instruction did not benefit low achieving poor readers. The overall effect size was close to zero. Possible reasons can be identified. LA readers' difficulties may have arisen from other sources such as lack of fluency, poor vocabulary, or poor reading comprehension. Another possibility is that the phonics instruction they received was inadequate. Inspection of the studies with LA readers revealed that only one study provided tutoring whereas seven involved classroom instruction. A third reason is low reliability and lack of power. Only eight comparisons contributed to the effect size.

Comprehending text successfully requires being able to read most of the words. Phonics programs teach children the alphabetic system and how to apply it to read words both in and out of text. As a result, phonics instruction should improve text reading as well as word reading. Findings of the meta-analysis confirmed that for beginners (kindergartners and first graders), phonics instruction benefited reading comprehension as much as it benefited reading miscellaneous words and decoding pseudowords. Also, phonics instruction benefited reading comprehension in older students

with RD. These results confirm the contribution of phonics instruction to text reading.

Systematic phonics instruction was found to boost spelling skills in younger but not older students. One factor contributing to younger children's spelling was the use of a scoring system that gave credit not only for correct spellings of words but also for letter–sound spellings that were phonetically plausible but not necessarily correct. These findings indicate that phonics instruction helps kindergartners and first graders acquire the alphabetic knowledge they need to begin learning to spell.

Phonics instruction failed to boost spelling among readers above first grade. Interestingly, a similar finding was detected in the NRP meta-analysis of phonemic awareness instruction that was found to enhance spelling ability significantly among younger children but not among older poor readers (Ehri et al., 2001). Also Swanson (2000) found reduced effects on spelling in a meta-analysis of instructional effects involving students with LD. One possible explanation is that poor readers experience special difficulty learning to spell (Bruck, 1993). Remediation of this difficulty may require specific instruction designed to teach spelling. Another explanation may be that as readers move up in the grades, success in spelling, at least in English, requires more than the skills taught in phonics programs. It requires remembering correct spellings of individual words. This involves teaching students to apply their knowledge of letter–sound correspondences, spelling patterns, and morphographic roots and affixes in order to detect regularities in the spellings of specific words (Ehri, 1997).

The effectiveness of two types of approaches for teaching phonics systematically was examined in our analysis, a synthetic approach teaching students to decode grapheme–phoneme units, and a larger-unit approach teaching students to blend subsyllablic units such as onsets, rimes, and phonograms. Two possible advantages of the larger-unit approach over the synthetic approach are that fewer sounds need to be blended when working with syllabic units than with graphophonemic units, and fewer schwa vowels attached to stop consonants have to be deleted when blending syllabic than graphophonemic units. These advantages suggest that larger-unit approaches might reveal larger effect sizes. However, findings indicated that the two approaches did not differ in their impact on reading, with both producing effects close to moderate in size.

When seven specific programs to teach systematic phonics were compared, they were found not to differ statistically in their effectiveness, with all producing a significant advantage in reading. Results showed that phonics instruction was effectively taught to individual students, small groups, and classrooms. Effect sizes were greater than zero and comparable in all three cases. Thus, these findings suggest that no one program or delivery system is better than others for teaching phonics systematically and that multiple ways can provide effective phonics instruction. However, the numbers of comparisons contributing to effect sizes were less than 10 in cases involving tutoring and specific programs. Reduced statistical power points to the need for more research before firm conclusions are drawn.

Although there is much interest in determining whether specific features of phonics programs influence their effectiveness, for example, the use of decodable text, it was not possible to examine these features because authors did not provide sufficient information to code the studies. We recommend that in the future, journal editors require researchers to include more details about instructional methods so that such an analysis becomes possible.

Over the years, several approaches to reading instruction have been developed that do not teach phonics systematically or that delay the introduction of phonics until children have learned to read whole words. The experiments in our database administered several types of nonsystematic or no-phonics instruction to control groups to evaluate the effectiveness of

phonics instruction: basal programs, regular curriculum, whole language approaches, whole word programs, and miscellaneous programs. Results showed that systematic phonics produced better reading than every type of program taught to control groups. The effect sizes were all positive favoring systematic phonics and were all statistically greater than zero. The fact that some forms of instruction taught to control groups included phonics means that the effect sizes we observed are actually underestimations, and effects of phonics instruction are actually stronger.

Much controversy exists about the relative effectiveness of phonics and whole language programs for helping beginners learn to read (Adams, 1990; Goodman, 1993; Grundin, 1994; McKenna, Stahl, & Reinking, 1994; Stahl, 1999; Taylor, 1998; Weaver, 1998). Some of the studies in our database examined the effectiveness of enriching whole language instruction with systematic phonics. Results were positive and suggest the importance of integrating systematic phonics instruction into whole language approaches rather than eliminating whole language from beginning reading instruction.

Comparison of Phonics and Phonemic Awareness Meta-Analysis

The NRP also conducted a similar meta-analysis of phonemic awareness (PA) instructional effects on learning to read and spell (Ehri et al., 2001). PA instruction differs from phonics instruction in that it is focused on teaching students to analyze and manipulate phonemes in the pronunciations of words. Manipulation may be taught by showing students how to use letters to manipulate the sounds they represent in words. However, instruction does not go beyond this to include practice reading or writing text. The overall effect size on reading was slightly greater in the PA analysis (d = 0.53) than in the phonics analysis (d = 0.41), but both were in the moderate range, indicating that the two alphabetic methods contribute substantially in helping children learn to read.

Methodological Strengths and Weaknesses of the Studies

We examined whether the largest effect sizes arose from well designed or weakly designed experiments. Three design characteristics were coded: the use of random assignment versus nonequivalent groups, presence versus absence of pretreatment differences between the phonics and control groups, and large versus small sample sizes.

Our studies varied in methodological rigor. Random assignment was used in 37% of the studies. Use of nonequivalent groups may have been more common because researchers often do not have the luxury of random assignment in school-based research. Sample size showed huge variation, with comparisons ranging from 20 to 320 students. Neither random assignment nor sample size differentially influenced mean effect sizes. Thus, the more rigorous designs did not yield different effect sizes from the less rigorous designs.

However, the design feature involving pretreatment differences did influence the effect size. A small number of comparisons (N = 5) involved groups that differed on pretests, with the phonics group displaying higher pretest scores than the control group, either on reading scores or IQ scores, in all five cases. Whereas the mean effect size on outcomes for these comparisons was close to zero, the mean effect size was positive and moderate in comparisons where the groups were equivalent on pretests. This suggests that weaker designs may have hindered the detection of effects. A similar finding was uncovered in the NRP meta-analysis of phonemic awareness instruction (Ehri et al., 2001) where studies with the strongest designs were found to display the largest effect sizes. These findings suggest that if there is a difference to be found, better

designed studies offer a greater chance of detecting it than poorly designed studies.

One common weakness of the studies was failing to provide needed information. Many studies did not fully describe the features included as part of phonics instruction. Also information about the full program used to teach reading was often missing. This precluded efforts to assess the importance of a comprehensive, balanced program that included not only systematic phonics but also other types of instruction thought to be important for learning to read. More research is needed to assess the contribution of various elements of phonics programs and to assess the impact of larger instructional contexts.

One potential criticism of our analysis is that we only considered published studies. Because negative findings are less apt to be published, the concern is that our pool of studies is biased and unrepresentative of a population of mostly unpublished studies finding no effects. We think this is unlikely. We uncovered 43 comparisons showing effect sizes of $d = 0.20$ or greater. In order for these to be statistical exceptions equaling the 5% expected by chance, there would have to be 860 comparisons showing effect sizes below 0.20. In a meta-analysis of instructional studies involving students with LD, Swanson and Hoskyn (1998) found that effect sizes on a composite cognitive/language measure were significantly larger for published studies than for unpublished studies. However, both effect sizes were statistically greater than zero, indicating that unpublished studies were not lacking in effects.

Implications for Classroom Instruction and Unanswered Questions

One of the primary questions that Congress asked the NRP to address was whether enough is known about systematic phonics instruction to make recommendations for classroom implementation. The answer is "Yes." Our findings came from studies conducted in many classrooms with typical teachers and typical students from a variety of backgrounds and SES levels. Most of the studies were published in the last 10 years so results are indicative of what can be accomplished when systematic phonics programs are implemented in today's classrooms. Our findings are consistent with other reports published earlier showing the positive results of systematic phonics instruction over a long period of time (Adams, 1990; Anderson et al., 1985; Chall, 1967, 1983, 1996a; Dykstra, 1968). These facts should persuade educators and the public that systematic phonics instruction is a valuable part of a successful classroom reading program. Our findings serve to illuminate the conditions that may make phonics instruction especially effective.

We were not able to answer many practical questions. One involved how long phonics instruction should continue through the grades for normally achieving readers. A few studies showed that when phonics instruction began in kindergarten and continued into second grade, effects on learning to read were sizeable and persisted. A few studies showed that small effects were evident in normally progressing readers in Grades 2 through 6 when phonics was introduced in these grades. However, more studies are needed. We have yet to determine whether phonics instruction is beneficial when it is continued beyond second grade for students who began learning to read with phonics.

Systematic phonics programs vary in many respects. Because instructional time is limited, teachers and publishers need to know which are the "active ingredients" yielding the most benefit. One likely ingredient involves the content covered. It is clear that major letter–sound correspondences, including short and long vowels and digraphs, need to be taught. However, written English has other regularities as well. It is unclear how many regularities should be taught explicitly. Should

children be taught to state regularities as well as how to apply them in reading and writing activities? More research is needed.

Another potential active ingredient is the use of decodable texts. Some systematic phonics programs provide books that are written carefully to focus mainly on the letter–sound relations that children have been taught. The intent is to help children develop skill and experience success in reading text as early as possible. The stories in such books often involve pigs doing jigs or cats in hats. Other systematic phonics programs make little or no use of decodable books that are rejected as too stilted and boring. Surprisingly, very little research has examined whether the use of decodable books has any influence on the progress that children make in learning to read. Such research is needed.

The motivational characteristics of phonics programs constitute another potential active ingredient to be investigated. Systematic phonics instruction has often been portrayed as involving "dull drill" and "meaningless worksheets." Whereas some of the approaches included in the meta-analysis may fit this characterization, others clearly do not. Few if any studies have investigated the importance of the motivational qualities of phonics programs. It seems self-evident that the specific techniques and activities used to teach phonics need to be relevant, motivating, and interesting in order to hold children's attention and to promote optimal learning. Moreover, when teaching is not only effective but also engaging and enjoyable, it seems likely that teachers will be more enthusiastic and committed to delivering instruction. Research is needed to identify the types of systematic phonics programs that are most motivating and to study the impact of various motivational features on programs' effectiveness.

Other potentially important questions include the following: (a) What knowledge about oral and written language do teachers need to have in order to teach systematic phonics effectively? (b) What types of preservice and inservice teacher education are required to prepare teachers to select and implement appropriate phonics approaches? (c) How flexible should programs be to accommodate variability among children? (d) How effective are computer-based phonics programs? and (e) Might metacognitive strategy instruction be incorporated into phonics programs to help students self-monitor and self-regulate the application of their alphabetic skills when they read independently?

There is currently much interest in whether systematic phonics instruction is effective for children who are learning English as a second language (ELL). However, most of our studies either provided no information about this population or intentionally excluded these students from the sample. Results of only one study pertained to ELL students, that by Stuart (1999) who included 86% ELL in her sample. The effect size she observed was large, indicating that phonics instruction helps ELL kindergartners learn to read more effectively than a whole language approach. However, more research is needed to replicate and extend this finding in order to identify the important ingredients of beginning reading instruction for ELL students.

Finally, it is important to underscore the place of phonics in a beginning reading program. Systematic phonics instruction by itself does not help students acquire all the processes they need to become successful readers. Phonics needs to be combined with other essential instructional components to create a complete and balanced reading program. Other sections of the NRP (2000) report indicated the importance of instruction to teach fluency, vocabulary, and reading comprehension strategies. In a meta-analysis of instructional studies employed with students having LD, Swanson (2000) observed significantly larger effect sizes on reading outcomes when direct skills instruction was combined with comprehension strategy instruction than when each was administered

separately to students. By emphasizing all the processes that contribute to growth in reading, teachers will have the best chance of making every child a reader.

APPENDIX A

Descriptions of the Specific Phonics Programs Examined in the Meta-Analysis

1. *Direct Instruction.* The Direct Instruction program is based on a behavioral analysis of the steps involved in learning to decode (Carnine & Silbert, 1979; Engelmann, 1980; Engelmann & Bruner, 1969, 1978, 1988; Engelmann & Osborn, 1987; Kameenui, Simmons, Chard, & Dickson, 1997). At the beginning of the program, students are not taught letter names but only letter–sound relations through highly structured instruction that uses cueing and reinforcement procedures derived from a behavioral analysis of instruction. The task of decoding is broken down into its component parts, and each of these parts is taught separately, from letter sounds to blending to reading words in context. Instruction is scripted and the lessons are fast paced, with high student participation. The text for the first-year program is written in a script that, although it preserves English spelling, contains printed marks that cue the reader about how the alphabetic system works, for example, silent letters, long vowels, and digraphs. Children practice in specially constructed books containing taught sounds, although children may be encouraged to read widely in children's literature as well (e.g., Meyer, 1983).

2. *Lovett Direct Instruction.* The synthetic phonics program used by Lovett and Steinbach (1997) and Lovett et al. (2000) adopts the Direct Instruction model to remediate the decoding and phonemic awareness difficulties of severely disabled readers. Children are taught phonological analysis and blending orally plus letter–sound associations in the context of word recognition and decoding instruction. Cumulative, systematic review and many opportunities for overlearning are hallmarks of this approach. New material is not introduced until the child fully masters previously instructed material.

3. *Lovett Analogy.* A second program also used with severely disabled readers by Lovett and Steinbach (1997) and Lovett et al. (2000) was adapted from the Benchmark Word Identification/Vocabulary Development program developed by Gaskins et al. (1988). This program is strongly metacognitive in its focus. It teaches children how to use four metacognitive strategies to decode words: reading words by analogy, detecting parts of words that are known, varying the pronunciation of vowels to maintain flexibility in decoding attempts, and "peeling off" prefixes and suffixes in words. Children learn a set of 120 key words exemplifying high-frequency spelling patterns, five words per day. They learn to segment the words into subunits so that they can use these known words and their parts to read other similarly spelled words. They learn letter–sound associations for vowels and affixes. Various types of texts provide children with practice applying the strategies taught.

4. *Lippincott.* The Lippincott Basic Reading Series (McCraken & Walcutt, 1975) is a direct code method that, from the outset, approaches reading from a phonic/linguistic perspective. Beginning with children's spoken language, the Lippincott program teaches in a systematic manner how to use the alphabetic code to move from printed words to oral language. Instruction begins with short *a* and builds knowledge of regular sound/symbol relationships. Children are first taught to decode phonetically regular words, with blending of phonic elements directly taught. Once they are proficient, long vowels and irregular spellings are introduced. Although the primary focus is on decoding, another goal of this method is the instant recognition of words. However,

rather than relying on a "context clue" approach to word recognition, children are taught how and why the letters come to represent these words, and they learn to "break the code" to decipher new words independently. Review and reinforcement are an integral part of the program. Spelling is sometimes taught as one component of the reading lesson with spelling lists developed from the words introduced in each unit of reading instruction.

5. *NRS by Beck and Mitroff.* The New Primary Grades Reading System for an Individualized Classroom (NRS) was developed by Beck and Mitroff (1972). It is a code-breaking approach. The program begins by teaching self-management skills, letter–sound correspondences, and chain blending to decode words. Children are taught to pronounce the first letter of a word followed by the second letter and then to blend the two sounds; then they pronounce the third letter and add it to the blend. In the first lesson, children are taught five isolated letter–sound relations, and once they are known, children are immediately taught to blend them to form real words. Subsequent letter–sounds are taught one at a time and blended with the earlier letters. Not only synthetic phonics but also analytic phonics is taught as children explore words and their parts. The method is linguistic as well, because the major spelling patterns of words are displayed in texts to draw attention to similarities and contrasts and because there is minimum teaching of explicit pronunciation rules. Instruction is individualized. After the first two levels, children work through the curriculum at different rates.

6. *Orton-Gillingham.* The Orton-Gillingham approach (Cox, 1991; Gillingham & Stillman, 1979) begins with the direct teaching of individual letters paired with their sounds using a visual-auditory-kinesthetic-tactile (VAKT) procedure that involves tracing the letter while saying its name and sound, blending letters together to read words and sentences, and finally reading short stories constructed to contain only taught sounds. Spelling words from dictation is also part of an Orton-Gillingham lesson. Each letter–sound is learned to mastery through repetition. More advanced lessons involve teaching learners to blend syllables together and read more complex texts. Among those approaches based on Orton and Gillingham's work are the Slingerland approach (Lovitt & DeMier, 1984), the Spaulding Approach, Recipe for Reading, and Alphabetic Phonics (Ogden, Hindman, & Turner, 1989). There are differences among these approaches, largely in the sequencing of materials, but they all have the general characteristics discussed.

7. *Sing, Spell, Read & Write.* The Sing, Spell, Read and Write program (Dickson, 1972) also teaches synthetic phonics. It consists of several charts, books (both readers and workbooks), letter and word cards, tests, and audio tapes. The tapes contain songs about several phonics generalizations. Through the tapes, the students learn the sounds of letters and letter combinations. Also songs combined with charts help students learn the spellings of words. The lessons begin by teaching letter sounds in isolation for each letter of the alphabet. When students have mastered certain sounds, they begin reading phonetic storybooks. The first five books each focus on a different vowel sound. The remaining books expand the vocabulary in a way that is consistent with the letter sounds taught. Students are taught to spell the words they learn to read, with the words presented in sentences. Most of the writing students do involves filling in blanks or answering questions related to words being learned. The program has a "racetrack" which is posted in classrooms and notes students' progress by placement of a race car on the chart.

Appendix B

Studies in the Phonics Database, Their Characteristics and Effect Sizes

Author and Year, Treatment	Characteristics of Training: Type of Phonics	Control Group	Tr. Unit	Length of Training	Characteristics of Part.: Grade/ Age	Reading Ability	SES
Blachman, Tangel, Ball, Black, & McGraw, 1999							
Blachman PA	Syn	Basal	SmG	2–3 yrs (41s,20m/d)	K	AR	Low
Blachman PA				(1st gr = 30 m/d)			
Blachman PA				(2nd gr = 30 m/d)			
Bond et al., 1995–1996							
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr. (20 lessons)	K	Nor	Var
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr.	1st	Nor	Var
Sing, Spell, Read, Write	Syn	Basal	Class	1 yr.	2nd	Nor	Var
Brown & Felton, 1990							
Lippincott	Syn	Wh.W.	SmG	2 yrs.	1st	AR	NG
Lippincott							
Eldredge, 1991							
Modified Whole Language	Syn	Basal	Class	1 yr. (15 m/d)	1st	AR	Low
Evans & Carr, 1985							
Traditional Basal	Misc	Wh.L.	Class	1 yr.	1st	Nor	Var
Foorman et al., 1991							
Synthetic basal	Syn	Wh.W.	Class	1 yr. (45 m/d)	1st	Nor	Mid
Foorman et al., 1997							
Orton-Gillingham	Syn	Wh.W.	SmG	1 yr. (60 m/d)	gr 2–3	RD	Mid
Onset-rime	LU	Wh.W.	SmG	1 yr.	gr 2–3	RD	Mid
Foorman et al., 1998							
Open Court	Syn	Wh.L.	Class	1 yr. (30 m/d)	1st	AR	Var
Embedded	LU	Wh.L.	Class	1 yr.	1st	AR	Var
Open Court	Syn	Wh.L.	Class	1 yr.	2nd	LA	Var
Embedded	LU	Wh.L.	Class	1 yr.	2nd	LA	Var
Freppon, 1991							
Sequential phonics	Misc	Wh.L.	Class	1 yr.	1st	Nor	Mid
Fulwiler & Groff, 1980							
Lippincott	Syn	Wh.W.	Class	1 yr.	1st	Nor	NG
Gersten, Darch, & Gleason, 1988							
Direct Instruction	Syn	Rg.cls.	Class	4 yrs.	K	AR	Low
Direct Instruction	Syn	Rg.cls.	Class	3 yrs.	1st	AR	Low
Gittelman & Feingold, 1983							
Intersensory Method	Syn	Misc.	Tutor	18 wks. (54s)	7–13 yrs.	RD	Mid
Greaney, Tunmer, & Chapman, 1997							
RRD-Rime analogy	LU	Wh.W.	Tutor	11 wks (31s, 30 m)	gr 2–5	LA	NG
RRD-Rime analogy							
Griffith, Klesius, & Kromey, 1992							
Traditional Basal	Misc	Wh.L.	Class	1 yr.	1st	Nor	NG
Haskell, Foorman, & Swank, 1992							
Analyze Onset-Rimes	Misc	Wh.W.	SmG	6 wks (15s, 20 m)	1st	Nor	Mid
Analyze Phonemes	Misc	Wh.W.	SmG	6 wks (15s, 20 m)	1st	Nor	Mid

Appendix B, continued

Features of Design				Effect Sizes on Posttests							
Group Assign.	Sig Pre-Test Diff	Total *N*	Time of Posttest	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
NE	No	159	Imm.	0.72	-0.17	1.08	0.94	.	1.04	.	.
		128	2nd yr tr.	0.64	0.35	0.81	0.53	.	0.86	.	.
		106	3rd yr tr.	0.36	0.42	0.55	0	.	0.45	.	.
NE	No	144	Imm.	0.51	0.38	.	.	.	1.01	0.13	.
NE	No	276	Imm.	0.25	0.23	.	0.14	.	0.6	0.03	.
NE	No	320	Imm.	0.38	0.44	.	0.18	.	0.55	0.33	.
R	No	47	Imm.	0.48	0.02	.	0.51	.	0.92	.	.
			2nd yr tr.	0.52	0.51	0.63	0.38	.	0.55	.	.
NE	No	105	Imm.	0.63	.	.	.	0.83	0.43	.	.
NE	NG	20* (*N*=247)	Imm.	0.6	.	.	.	0.6	.	.	.
NE	No	6* (*N*=80)	Imm.	2.27	1.92	2.67	2.21	.	.	.	.
NG	Yes	67	Imm.	0.27	0.17	0.58	0.05	.	.	.	.
NG	No	85	Imm.	-0.11	-0.19	0.09	-0.23	.	.	.	.
NE	No	68	Imm.	0.91	1.63	1.14	0.56	0.32	.	.	.
NE	No	70	Imm.	0.36	0.56	0.51	0.26	0.1	.	.	.
NE	No	35	Imm.	0.12	0.52	0.32	-0.19	-0.19	.	.	.
NE	No	57	Imm.	0.03	0.37	0.22	-0.25	-0.24	.	.	.
NE	Yes	24	Imm.	0	.	.	.	.	.	0	.
NE	NG	147	Imm.	0.84	.	0.91	.	0.76	.	.	.
NE	No	101	4th yr tr.	0.24	.	.	0.16	0.28	.	.	0.27
NE	No	141	3rd yr tr.	0	.	.	-0.12	0.11	.	.	0.02
R	No	56	Imm.	0.53	0.76	0.67	0.12	0.57	.	.	.
R	No	36	Imm.	0.37	0.39	.	.	.	0.51	0.2	.
		34	follow up	0.56	0.47	.	.	.	0.76	0.44	.
NE	No	24	Imm.	-0.33	-1.11	.	-0.54	-0.43	0.78	.	.
R	No	24	Imm.	0.14	0.2	0.09	.	.	.	.	.
R	No	24	Imm.	-0.07	-0.08	-0.06	.	.	.	.	.

Appendix B, continued

Author and Year, Treatment	Characteristics of Training: Type of Phonics	Control Group	Tr. Unit	Length of Training	Characteristics of Part.: Grade/ Age	Reading Ability	SES
Klesius, Griffith, Zielonka, 1991							
Traditional Basal	Misc	Wh.L.	Class	1 yr.	1st	Nor	Var
Leach & Siddall, 1990							
Direct Instruction	Syn	Misc.	Tutor	10 wks. (15 m/d)	1st	Nor	NG
Leinhardt & Engel, 1981							
NRS-study 2 (Beck)	Syn	Basal	SmG	1 yr.	1st	Nor	NG
NRS-study 3 (Beck)	Syn	Basal	SmG	1 yr.	1st	Nor	NG
NRS-study 4 (Beck)	Syn	Basal	SmG	1 yr.	1st	Nor	NG
NRS-study 6 (Beck)	Syn	Basal	SmG	1 yr.	1st	Nor	NG
Lovett, Ransby, Hardwick, Johns, & Donaldson, 1989							
Decoding Skills	Syn	Misc.	SmG	40 ses (33–40h)	8–13 yr.	RD	Mid
Lovett & Steinbach, 1997							
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 2/3	RD	NG
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 4	RD	NG
Lovett Analogy	LU	Misc.	SmG	9wks (35h)	gr 5/6	RD	NG
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 2/3	RD	NG
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 4	RD	NG
Lovett Direct Instruction	Syn	Misc.	SmG	9wks (35h)	gr 5/6	RD	NG
Lovett, Warren-Chaplin, Ransby, & Borden, 1990							
Analytic	Misc	Misc.	SmG	9wks (35h)	7–13 yr	RD	Mid
Lovett, et al., (2000)							
Dir. Instruction + Analogy	Com	Misc.	SmG	70h	6–13 yr	RD	Var
Analogy + Direct Instruction	Com	Misc.	SmG	70h	6–13 yr	RD	Var
Lovett Direct Instruction	Syn	Misc.	SmG	70h	6–13 yr	RD	Var
Lovett Analogy	LU	Misc.	SmG	70h	6–13 yr	RD	Var
Lum & Morton, 1984							
Spelling Mastery	Misc	Rg.cls.	Class	1 yr. (20–30 m/d)	2nd	Nor	NG
Mantzicopoulos, Morrison, Stone, & Setrakian, 1992							
Phonetic read/spell	Misc	Rg.cls.	Tutor	50s (1h/wk)	1st	AR	Mid
Phonetic read/spell							
Marston, Deno, Kim, Diment, & Rogers, 1995							
Direct Instruction	Syn	Rg.cls.	Class	10 wks (45 m/d)	gr 1–6	LA	NG
Martinussen & Kirby, 1998							
Succesive phonics	Syn	Rg.cls.	SmG	8 wks (40–60 m/wk)	K	AR	NG
Oakland, Black, Stanford, Nussbaum, & Balise, 1998							
Orton-Gillingham	Syn	Rg.cls.	SmG	2 yrs.(350h)	11 yrs.	RD	NG
Santa & Hoien, 1999							
RRD-Early Steps	LU	Wh.L.	Tutor	1 yr.(30 m/d)	1st	AR	Var
RRD-Early Steps							

Appendix B, continued

Features of Design				Effect Sizes on Posttests							
Group Assign.	Sig Pre-Test Diff	Total *N*	Time of Posttest	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
NE	Yes	6* (*N*=112)	Imm.	0.2	.	.	0.36	0.18	0.07		
R	No	20	Imm.	1.99	.	.	.	1.8	.	2.18	.
NE	Y/Adj	187	Imm.	0.45	0.45	.	.	.	.	.	.
NE	Y/Adj	263	Imm.	0.44	0.44	.	.	.	.	.	.
NE	Y/Adj	256	Imm.	0.33	0.33	.	.	.	.	.	.
NE	Y/Adj	241	Imm.	0.7	0.7	.	.	.	.	.	.
R	No	118	Imm.	0.39	0.78	0.7	0.42	0.07	0.1	0.27	.
R	No	28	Imm.	0.49	-0.12	0.85	.	.	0.75	.	.
R	No	22	Imm.	1.41	0.84	2.06	.	.	1.33	.	.
R	No	24	Imm.	-0.25	-0.49	-0.15	.	.	-0.1	.	.
R	No	32	Imm.	0.24	0.02	0.24	.	.	0.46	.	.
R	No	25	Imm.	1.42	1.03	1.53	.	.	1.7	.	.
R	No	27	Imm.	0.09	-0.24	0.25	.	.	0.25	.	.
R	No	36	Imm.	0.16	0.13	0.11	0.23	.	.	.	.
R	No	37	Imm.	0.6	0.36	1	0.15	0.27	1.22	.	.
R	No	32	Imm.	0.21	0.04	0.55	-0.2	0.12	0.52	.	.
R	No	40	Imm.	0.24	0.21	0.36	-0.19	0.42	0.42	.	.
R	No	42	Imm.	0.5	0.47	0.75	0.01	0.6	0.66	.	.
NE	No	36	Imm.	0.38	0.31	.	0.45	.	.	.	.
R	No	112	Imm.	0.53	.	.	.	.	0.53	.	.
		112	follow up	0.32	.	0.33	0.3	0.08	0.56	.	.
NE	Yes	53	Imm.	0.01	.	.	.	.	.	0.01	.
R	No	26	Imm.	0.62	0.53	0.63	0.68	.	0.62	.	.
NE	Yes	48	2nd yr tr.	0.54	0.71	.	0.23	0.62	0.61	.	.
NE	No	49	Imm.	0.76	0.93	.	0.63	0.73	.	.	.
		41	follow up	0.86	0.57	.	.	0.87	1.15	.	.

Appendix B, continued

Author and Year, Treatment	Characteristics of Training: Type of Phonics	Control Group	Tr. Unit	Length of Training	Characteristics of Part.: Grade/ Age	Reading Ability	SES
Silberberg, Iversen, & Goins, 1973							
Lippincott	Syn	Wh.W.	SmG	1 yr.	gr 3	RD	NG
Orton-Gillingham	Syn	Wh.W.	SmG	1 yr.	gr 3	RD	NG
Lippincott							
Orton-Gillingham							
Snider, 1990							
Direct Instruction	Syn	Basal	SmG	1 yr. (60 m/d)	1st	Nor	Mid
Stuart, 1999							
Jolly Phonics	Syn	Wh.L.	Class	12 wks (60 m/d)	K	AR	Low
Jolly Phonics							
Torgesen et al., 1999							
Lindamood PA	Syn	Rg.cls.	Tutor	2.5 yrs. (80 m/wk)	K	AR	NG
Embedded	LU	Rg.cls.	Tutor	2.5 yrs. (80 m/wk)	K	AR	NG
Lindamood PA							
Embedded							
Lindamood PA							
Embedded							
Traweek & Berninger, 1997							
Direct Instruction	Syn	Wh.L.	Class	1 yr.	1st	AR	Low
Tunmer & Hoover, 1993							
RRD-Phonograms	LU	Rg.cls.	Tutor	42 s (30 m/d)	1st	AR	NG
Umbach, Darch, & Halpin, 1989							
Direct Instruction	Syn	Basal	SmG	1 yr. (50 m/d)	1st	AR	Low
Vandervelden & Siegel, 1997							
Developmental	Misc	Rg.cls.	SmG	12 wks (30–45 m/wk)	K	AR	Low
Vickery, Reynolds, & Cochran, 1987							
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	3rd	Nor	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	4th	Nor	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	5th	Nor	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	6th	Nor	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	3rd	LA	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	4th	LA	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	5th	LA	NG
Orton-Gillingham	Syn	Rg.cls.	Class	1 yr. (55 m/d)	6th	LA	NG
Wilson & Norman, 1998							
Sequential phonics	Syn	Wh.L.	Class	1 yr.	2nd	Nor	NG

Abbreviations:
* = class was used as the unit of analysis
AR = At Risk
Com = Combination
Comp = Comprehension
Dec = Decoding
Gen. Read = General reading
gr = grade
h = hour
h/wk = hours per week
Imm. = Immediate
K = Kindergarten
LA = Low Achievement
LU = Larger Units
M = mean
m = minutes
m/d = minutes per day
m/wk = minutes per week
Mid = Middle Class
Misc = Miscellaneous
N = Number of participants
NE = Non Equivalent groups
NG = Not Given
Nonw = Nonword reading
Nor = Normal Readers
Oral Read = Oral reading
R = Random assignment

Appendix B, continued

Features of Design				Effect Sizes on Posttests							
Group Assign.	Sig Pre-Test Diff	Total *N*	Time of Posttest	Mean	Word ID	Dec	Spell	Comp	Nonw	Oral Read	Gen. Read
NE	No	69	Imm.	0.5	0.7	.	.	0.36	.	0.45	.
NE	Yes	65	Imm.	0.04	0.31	.	.	0.09	.	-0.29	.
		62	follow up	0.33	0.37	.	.	-0.04	.	0.66	.
		58	follow up	-0.47	-0.19	.	.	-0.81	.	-0.4	.
NE	No	66	follow up	0.38	.	0.6	0.44	0.1	.	.	.
NE	Y/Adj	112	Imm.	0.73	0.56	.	1.11	0.36	0.9	.	.
		112	follow up	0.28	0.11	.	0.5	0.31	-0.03	0.49	.
R	No	65	Imm.	0.33	0.08	.	.	.	0.58	.	.
R	No	68	Imm.	0.32	0.52	.	.	.	0.12	.	.
		65	2nd yr tr.	0.75	0.64	.	.	0.49	1.13	.	.
		68	2nd yr tr.	0.28	0.24	.	.	0.29	0.31	.	.
		65	3rd yr tr.	0.67	0.67	.	0.64	0.36	1.01	.	.
		68	3rd yr tr.	0.17	0.25	.	0.1	0.17	0.16	.	.
NE	Y/Adj	38	Imm.	0.07	0.07	.	.	.	.	.	.
NG	No	64	Imm.	3.71	2.94	.	1.63	.	1.49	8.79	.
R	No	31	Imm.	1.19	1.3	.	.	1.08	.	.	.
NE	No	29	Imm.	0.47	0.04	.	1.11	.	0.57	0.15	.
NE	NG	63	Imm.	0.04	.	.	.	.	.	.	0.04
NE	NG	71	Imm.	0.04	.	.	.	.	.	.	0.04
NE	NG	74	Imm.	0.61	.	.	.	.	.	.	0.61
NE	NG	79	Imm.	0.43	.	.	.	.	.	.	0.43
NE	NG	46	Imm.	0.63	.	.	.	.	.	.	0.63
NE	NG	47	Imm.	0.19	.	.	.	.	.	.	0.19
NE	NG	45	Imm.	-0.2	.	.	.	.	.	.	-0.2
NE	NG	41	Imm.	0.13	.	.	.	.	.	.	0.13
NE	No	54	Imm.	-0.47	-0.33	.	.	-0.61	.	.	.

RD = Reading Disabled
Rg.cls. = Regular class
s = session(s)
SmG = Small group
Spell = Spelling
Syn = Synthetic
tr = training
Var = Varied
Word ID = Word Identification
Wh.L. = Whole Language
Wh.W. = Whole Word
wks = weeks
Y/Adj = Yes, but means were adjusted for pretest differences
yr. = year

Notes

[1] One exception occurred. Findings of the study by Tunmer and Hoover (1993) were reported also by Iversen and Tunmer (1993) and were included in the PA meta-analysis. However, the treatment–control comparisons used in the two meta-analyses involved different control groups. The phonics comparison used performance of the control group receiving the standard intervention provided by the school, whereas the PA comparison used performance of the control group whose instruction was identical to that of the treatment group except for the absence of tutoring in phonemic awareness.

[2] This report differs slightly from the NRP report. The number of comparisons showing pretreatment differences was seven in the NRP report but five here. The difference occurred because overall studies rather than individual treatment–control group comparisons within studies were coded for the presence of pretest differences in the NRP report. In this report, each treatment–control comparison was coded separately. Effect sizes remained almost identical despite the difference.

Acknowledgements

This research was conducted by the Alphabetics subgroup of the National Reading Panel and was supported by the National Institute of Child Health and Human Development. Gratitude is expressed to several people for their contributions: Blair Johnson and Harris Cooper, who provided statistical advice; Vinita Chhabra, who conducted the electronic searches and located studies from libraries; Barbara Schuster, who supervised the coders; Alisa Kenny, Maggie Toplak, Michelle Eidlitz, Marjolaine Limbos, and Khalil Nourani, who coded the studies; Joan Pagnucco, who entered the coded information into the database; Barbara Foorman, Tim Shanahan, and Joanne Yatvin, who provided helpful suggestions on the manuscript.

References

Note. An asterisk marks those publications providing data for the meta-analysis.

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

Anderson, R. C., Hiebert, E. F., Wilkinson, I. A. G., & Scott, J. (1985). *Becoming a nation of readers*. Champaign, IL: Center for the Study of Reading.

Aukerman, R. (1971). *Approaches to beginning reading*. New York: Wiley.

Aukerman, R. (1981). *The basal reader approach to reading*. New York: Wiley.

Aukerman, R. (1984). *Approaches to beginning reading* (2nd ed.). New York: Wiley.

Balmuth, M. (1982). *The roots of phonics: A historical introduction*. New York: McGraw-Hill.

Barr, R., Kamil, M., Mosenthal, P., & Pearson, P. (1991). *Handbook of reading research* (Vol. 2). New York: Longman.

Beck, I., & Mitroff, D. (1972). *The rationale and design of a primary grades reading system for an individualized classroom*. Pittsburgh, PA: University of Pittsburgh Learning Research and Development Center.

Berninger, V., Abbott, R., Billingsley, F., & Nagy, W. (2001). Processes underlying timing and fluency of reading: Efficiency, automaticity, coordination and morphological awareness. In M. Wolf (Ed.), *Dyslexia, fluency and the brain*. Timonium, MD: York Press.

Blachman, B., Ball, E., Black, R., & Tangel, D. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? *Reading and Writing: An Interdisciplinary Journal, 6,* 1–18.

*Blachman, B., Tangel, D., Ball, E., Black, R., & McGraw, D. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal, 11,* 239–273.

*Bond, C., Ross, S., Smith, L., & Nunnery, J. (1995–96). The effects of the sing, spell, read, and write program on reading achievement of beginning readers. *Reading Research and Instruction, 35,* 122–141.

*Brown, I., & Felton, R. (1990). Effects of instruction on beginning reading skills in children at risk for reading disability. *Reading and Writing: An Interdisciplinary Journal, 2,* 223–241.

Bruck, M. (1993). Component spelling skills of college students with childhood diagnoses of dyslexia. *Learning Disability Quarterly, 16,* 171–184.

Carnine, D., & Silbert, J. (1979). *Direct Instruction reading.* Columbus, OH: Merrill.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Chall, J. S. (1983). *Learning to read: The great debate* (2nd ed.). New York: McGraw-Hill.

Chall, J. (1996a). *Learning to read: The great debate* (revised, with a new foreword). New York: McGraw-Hill.

Chall, J. (1996b). *Stages of reading development.* Fort Worth, TX: Harcourt Brace.

Clay, M. M. (1993). *Reading recovery: A guidebook for teachers in training.* Portsmouth, NH: Heinemann.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cox, A. (1991). *Structures and techniques: Multisensory teaching of basic language skills.* Cambridge, MA: Educators Publishing Service.

Dickson, S. (1972). *Sing, spell, read, and write.* Chesapeake, VA: International Learning Systems.

Dykstra, R. (1968). The effectiveness of code- and meaning-emphasis in beginning reading programs. *The Reading Teacher, 22,* 17–23.

Ehri, L. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 383–417). New York: Longman.

Ehri, L. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 323–358). Newark, DE: International Reading Association.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti & L. Rieben (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237–268). Mahwah, NJ: Erlbaum.

Ehri, L. C. (1998). grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

Ehri, L., Deffner, N., & Wilce, L. (1984). Pictorial mnemonics for phonics. *Journal of Educational Psychology, 76,* 880–893.

Ehri, L., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287.

*Eldredge, L. (1991). An experiment with a modified whole language approach in first-grade classrooms. *Reading Research and Instruction, 30,* 21–38.

Engelmann, S. (1980). *Direct Instruction.* Englewood Cliffs, NJ: Prentice Hall.

Engelmann, S., & Bruner, E. (1969). *DISTAR reading program.* Chicago: SRA.

Engelmann, S., & Bruner, E. (1978). *DISTAR reading: An instructional system.* Chicago: SRA.

Engelmann, S., & Bruner, E. (1988). *Reading Mastery I/II fast cycle: Teacher's guide.* Chicago: SRA.

Engelmann, S., & Osborn, J. (1987). *DISTAR Language I.* Chicago: SRA.

*Evans, M., & Carr, T. (1985). Cognitive abilities, conditions of learning, and the early development of reading skills. *Reading Research Quarterly, 20,* 327–350.

Flood, J., Jensen, J., Lapp, D., & Squire, J. (Eds.). (1991). *Handbook of research on teaching the English language arts.* New York: Macmillan.

*Foorman, B., Francis, D., Fletcher, J., & Schatschneider, C. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37–55.

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

*Foorman, B., Francis, D., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1,* 255–276.

*Freppon, P. (1991). Children's concepts of the nature and purpose of reading in different instructional settings. *Journal of Reading Behavior, 23,* 139–163.

*Fulwiler, G., & Groff, P. (1980). The effectiveness of intensive phonics. *Reading Horizons, 21,* 50–54.

Gaskins, I. W., Downer, M. A., Anderson, R. C., Cunningham, P. M., Gaskins, R. W., Schommer, M., & the Teachers of Benchmark School. (1988). A metacognitive approach to phonics: Using what you know to decode what you don't know. *Remedial and Special Education, 9,* 36–41.

*Gersten, R., Darch, C., & Gleason, M. (1988). Effectiveness of a Direct Instruction academic kindergarten for low-income students. *The Elementary School Journal, 89,* 227–240.

Gillingham, A., & Stillman, B. (1979). *Remedial training for children with specific disability in reading, spelling, and penmanship.* Cambridge, MA: Educators Publishing Service.

*Gittelman, R., & Feingold, I. (1983). Children with reading disorders—I. Efficacy of reading remediation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 24,* 167–191.

Goodman, K. S. (1993). *Phonics phacts.* Portsmouth, NH: Heinemann.

*Greaney, K., Tunmer, W., & Chapman, J. (1997). Effects of rime-based orthographic analogy training on the word recognition skills of children with reading disability. *Journal of Educational Psychology, 89,* 645–651.

*Griffith, P., Klesius, J., & Kromrey, J. (1992). The effect of phonemic awareness on the literacy development of first-grade children in a traditional or a whole language classroom. *Journal of Research in Childhood Education, 6,* 85–92.

Grundin, H. (1994). If it ain't whole, it ain't language—or back to the basics of freedom and dignity. In F. Lehr & J. Osborn (Eds.), *Reading, language, and literacy* (pp. 77–88). Mahwah, NJ: Erlbaum.

Harris, T., & Hodges, R. (Eds.). (1995). *The literacy dictionary*. Newark, DE: International Reading Association.

*Haskell, D., Foorman, B., & Swank, P. (1992). Effects of three orthographic/phonological units on first-grade reading. *Remedial and Special Education, 13,* 40–49.

Hiebert, E., Colt, J., Catto, S., & Gary, E. (1992). Reading and writing of first-grade students in a restructured chapter 1 program. *American Educational Research Journal, 29,* 545–572.

Holdaway, D. (1979). *The foundations of literacy*. Sydney, Australia: Ashton-Scholastic.

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

Johnson, B. (1989). DSTAT: *Software for the meta-analysis review of research literatures*. Hillsdale, NJ: Erlbaum.

Johnson, B., & Eagly, A. (2000). Quantitative synthesis of social psychological research. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 496–528). London: Cambridge University Press.

Kameenui, E. J., Simmons, D. C., Chard, D., & Dickson, S. (1997). Direct Instruction reading. In S. A. Stahl & D. A. Hayes (Eds.), *Instructional models in reading,* (pp. 59–84). Mahwah, NJ: Erlbaum.

*Klesius, J., Griffith, P., & Zielonka, P. (1991). A whole language and traditional instruction comparison: Overall effectiveness and development of the alphabetic principle. *Reading Research and Instruction, 30,* 47–61.

*Leach, D., & Siddall, S. (1990). Parental involvement in the teaching of reading: A comparison of hearing reading, paired reading, pause, prompt, praise, and direct instruction methods. *British Journal of Educational Psychology, 60,* 349–355.

*Leinhardt, G., & Engel, M. (1981). An iterative evaluation of NRS: Ripples in a pond. *Evaluation Review, 5,* 579–601.

Lindamood, C. H., & Lindamood, P. C. (1984). *Auditory discrimination in depth*. Austin, TX: PRO-ED, Inc.

Lippincott basic reading. (1981). Riverside, NJ: Macmillan.

Lloyd, S. (1993). *The phonics handbook*. UK: Jolly Learning.

*Lovett, M., & Steinbach, K. (1997). The effectiveness of remedial programs for reading disabled children of different ages: Does the benefit decrease for older children? *Learning Disability Quarterly, 20,* 189–210.

*Lovett, M., Warren-Chaplin, P., Ransby, M., & Borden, S. (1990). Training the word recognition skills of reading disabled children: Treatment and transfer effects. *Journal of Educational Psychology, 82,* 769–780.

*Lovett, M., Lacerenza, L., Borden, S., Frijters, J., Steinbach, K., & DePalma, M. (2000). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology, 92,* 263–283.

*Lovett, R., Ransby, M., Hardwick, N., Johns, M., & Donaldson, S. (1989). Can dyslexia be treated? Treatment-specific and generalized treatment effects in dyslexic children's response to remediation. *Brain and Language, 37,* 90–121.

Lovitt, T. C., & DeMier, D. M. (1984). An evaluation of the Slingerland method with LD youngsters. *Journal of Learning Disabilities, 17,* 267–272.

*Lum, T., & Morton, L. (1984). Direct instruction in spelling increases gain in spelling and reading skills. *Special Education in Canada, 58,* 41–45.

*Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the SEARCH/TEACH tutoring approach with middle-class students at risk for reading failure. *Elementary School Journal, 92,* 573–586.

*Marston, D., Deno, S., Kim, D., Diment, K., & Rogers, D. (1995). Comparison of reading intervention approaches for students with mild disabilities. *Exceptional Children, 62,* 20–37.

*Martinussen, R., & Kirby, J. (1998). Instruction in successive and phonological processing to improve the reading acquisition of at-risk kindergarten children. *Developmental Disabilities Bulletin, 26,* 19–39.

McCracken, G., & Walcutt, C. (1963). *Basic reading*. Philadelphia, PA: Lippincott.

McCracken, G., & Walcutt, C. (1975). *Basic reading*. Philadelphia, PA: Lippincott.

McKenna, M., Stahl, S., & Reinking, D. (1994). A critical commentary on research, politics, and whole language. *Journal of Reading Behavior, 26,* 211–233.

Meyer, L. A. (1983). Increased student achievement in reading: One district's strategies. *Research in Rural Education, 1,* 47–51.

Morris, D. (1992). *Case studies in teaching beginning readers: The Howard Street tutoring manual.* Boone, NC: Fieldstream.

Morris, D., & Perney, J. (1984). Developmental spelling as a predictor of first-grade reading achievement. *Elementary School Journal, 84,* 441–457.

National Reading Panel (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Rockville, MD: NICHD Clearinghouse.

*Oakland, T., Black, J., Stanford, G., Nussbaum, N., & Balise, R. (1998). An evaluation of the dyslexia training program: A multisensory method for promoting reading in students with reading disabilities. *Journal of Learning Disabilities, 31,* 140–147.

Ogden, S., Hindman, S., & Turner, S. (1989). Multisensory programs in the public schools: A brighter future for LD children. *Annals of Dyslexia, 39,* 247–267.

Open Court Reading. (1995). *Collections for young scholars*. Chicago: SRA/McGraw-Hill.

Pearson, P., Barr, R., Kamil, M., & Mosenthal, P. (1984). *Handbook of reading research*. New York: Longman.

Perfetti, C. (1992). The representation problem in reading acquisition. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

Purves, A. (Ed.). (1994). *Encyclopedia of English studies and language arts*. New York: Scholastic.

Rack, J., Snowling, M., & Olson, R. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 29–53.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Thousand Oaks, CA: Sage.

Routman, R. (1996). *Literacy at the crossroads*. Portsmouth, NH: Heinemann.

*Santa, C., & Hoien, T. (1999). An assessment of early steps: A program for early intervention of reading problems. *Reading Research Quarterly, 34,* 54–79.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by eye and by ear* (pp. 293–317). Cambridge, MA: MIT Press.

Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading achievement. *Journal of Educational Psychology, 76,* 1309–1324.

*Silberberg, N., Iversen, I., & Goins, J. (1973). Which remedial reading method works best? *Journal of Learning Disabilities, 6,* 18–27.

*Snider, V. (1990). Direct Instruction reading with average first graders. *Reading Improvement, 27,* 143–148.

Stahl, S. (1999). Why innovations come and go: The case of whole language. *Educational Researcher, 28,* 13.

Stahl, S., Duffy-Hester, A., & Stahl, K. (1998). Everything you wanted to know about phonics (But were afraid to ask). *Reading Research Quarterly, 35,* 338–355.

Stahl, S. A., & Miller, P. D. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. *Review of Educational Research, 59*(1), 87–116.

*Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology, 69,* 587–605.

Swanson, H. L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32,* 504–532.

Swanson, H. L. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In R. Gersten, E. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 1–30). Mahwah, NJ: Erlbaum.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68,* 277–321.

Tangel, D., & Blachman, B. (1995). Effect of phoneme awareness instruction on the invented spelling of first-grade children: A one-year follow-up. *Journal of Reading Behavior, 27,* 153–185.

Taylor, D. (1998). *Beginning to read & the spin doctors of science: The political campaign to change America's mind about how children learn to read.* Urbana, IL: National Council of Teachers of English.

*Torgesen, J., Wagner, R., Rashotte, C., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593.

*Traweek, K., & Berninger, V. (1997). Comparisons of beginning literacy programs: Alternative paths to the same learning outcome. *Learning Disability Quarterly, 20,* 160–168.

Tunmer, W., & Chapman, J. (1998). Language prediction skill, phonological recoding ability, and begin-

ning reading. In C. Hulme & R. Joshi (Eds.), *Reading and Spelling: Development and Disorders* (pp. 33–67). Mahwah, NJ: Erlbaum.

*Tunmer, W., & Hoover, W. (1993). Phonological recoding skill and beginning reading. *Reading and Writing: An Interdisciplinary Journal, 5,* 161–179.

*Umbach, B., Darch, C., & Halpin, G. (1989). Teaching reading to low performing first graders in rural schools: A comparison of two instructional approaches. *Journal of Instructional Psychology, 16,* 23–30.

*Vandervelden, M., & Siegel, L. (1997). Teaching phonological processing skills in early literacy: A developmental approach. *Learning Disability Quarterly, 20,* 63–81.

*Vickery, K., Reynolds, V., & Cochran, S. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia, 37,* 189–200.

Weaver, C. (1998). Experimental research: On phonemic awareness and on whole language. In C. Weaver (Ed.), *Reconsidering a balanced approach to reading* (pp. 321–371). Urbana, IL: National Council of Teachers of English.

Wendon, L. (1992). First steps in Letterland. Cambridge, UK: Letterland, Ltd.

*Wilson, K., & Norman, C. (1998). Differences in word recognition based on approach to reading instruction. *Alberta Journal of Educational Research, 44,* 221–230.

Authors

LINNEA C. EHRI is Distinguished Professor of Educational Psychology at the Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016; *LEhri@gc.cuny.edu*. Her research interests center on literacy and involve reading and spelling acquisition processes, reading instruction, and causes of difficulty.

SIMONE R. NUNES is a doctoral student in Educational Psychology at the Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016; *SNunes@gc.cuny.edu*. Her research interests include literacy acquisition and instruction, teacher research, and literacy acquisition across different languages.

STEVEN A. STAHL is Professor of Reading Education at the University of Georgia, 309 Aderhold Hall, Athens, GA 30602, and Co-Director of the Center for the Improvement of Early Reading Achievement. His research interests center on reading, including instruction of word recognition, phonological awareness, spelling, and vocabulary.

DALE M. WILLOWS is Professor of Psychology at the Ontario Institute for Studies in Education of the University of Toronto, 252 Bloor Street West, Toronto, Ontario, M5S 1V6, Canada; *dwillows@oise.utoronto.ca.* As a researcher and practitioner in the field of reading/writing, her current focus is on facilitating the implementation of research-based findings on literacy acquisition in school districts in Canada and the United States.