Reading Research Quarterly
Vol. 36, No. 3
July/August/September 2001

(pp. 250–287)

Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis

Linnea C. Ehri
City University of New York Graduate Center, New York, USA

Simone R. Nunes
Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Ministry of Education, Brasília, Brazil

Dale M. Willows
Barbara Valeska Schuster
Zohreh Yaghoub-Zadeh
Ontario Institute for Studies in Education, Toronto, Canada

Timothy Shanahan
University of Illinois, Chicago, USA

In 1997, the U.S. Congress directed that a national panel be convened to assess the status of research-based knowledge regarding the effectiveness of various approaches for teaching children to read. The Director of the National Institute of Child Health and Human Development in consultation with the Secretary of Education constituted the National Reading Panel (NRP), which was composed of 14 individuals. Members of the (NRP) Panel formed subgroups to locate research examining various categories of studies. One of the subgroups focused on alphabetics and conducted two meta-analyses, one on phonemic awareness instruction and one on phonics instruction. The final report was submitted to Congress in April 2000. Results of the phonemic awareness analysis are presented here.

When today's educators consider the ingredients of effective programs to teach children to read, phonemic awareness (abbreviated PA) receives much attention. The purpose of this study was to examine the scientific evidence supporting claims about phonemic awareness instruction. We sought answers to several questions: Is phonemic awareness instruction effective in helping children learn to read? Under what circumstances and for which children is it most effective? Were studies showing its effectiveness designed to yield scientifically valid findings? How applicable are these findings to classroom practice?

There were several reasons why the NRP selected phonemic awareness instruction for review and analysis. First, correlational studies have identified phonemic awareness and letter knowledge as the two best school-entry predictors of how well children will learn to read during the first 2 years of instruction (Share, Jorm, Maclean, & Matthews, 1984). Second, many experimental studies have been conducted to evaluate the effectiveness of PA instruction in facilitating reading acquisition. However, a recent critique appearing in *Reading Research Quarterly* suggested that some of the PA studies are seriously flawed (Troia, 1999). Third, there is currently much interest in PA programs among teachers, principals, publishers, and even legislators because they are claimed to be effective in improving children's success in learning to read.

To evaluate the adequacy and strength of the evidence, we conducted a meta-analysis. We located all the experimental studies that administered PA instruction to students, that included a control group receiving non-PA instruction or no special instruction, and that measured the impact of PA instruction on reading outcomes. We found 52 published studies that met our criteria. The studies varied in many respects. Different types of phonemic awareness were taught. The participants ranged from preschoolers to sixth graders and included students at risk for reading problems as well as students classified as reading disabled. The instruction was delivered by classroom teachers in some studies and by researchers or computers in other studies. Children were tutored individually, or they received instruction in small groups or larger classroom groups. The statistic we used was effect size, that is, the difference between mean performance of the treatment and control groups in standard deviation units. The meta-analytic procedure allowed us to examine not only whether PA instruction exerted a statistically significant impact on reading across all of these variations but also whether these variations influenced the size of the impact.

What is phonemic awareness? Phonemes are the smallest units comprising *spoken* language. English consists of about 41 phonemes. Phonemes combine to form syllables and words. A few words have only one phoneme, such as *a* or *oh*. Most words consist of a blend of phonemes, such as *go* with 2 phonemes, or *check* with 3 phonemes, or *stop* with 4 phonemes. In the text below, individual phonemes are represented with IPA (International Phonetic Alphabet) symbols between backslashes (e.g., /g/) to contrast them with letters represented by capitals (e.g., G). Phonemes are different from graphemes, which are units of written language and represent phonemes in the spellings of words (Venezky, 1970, 1999). Graphemes may consist of one letter, for example, P, T, K, A, N; or multiple letters, CH, SH, TH, -CK, EA, -IGH, each symbolizing one phoneme.

Phonemic awareness refers to the ability to focus on and manipulate phonemes in spoken words (Liberman, Shankweiler, Fischer, & Carter, 1974). Simply discriminating phonemes in words—for example, recognizing that *tan* sounds different from *Dan*—is not PA. PA is different from phonological awareness, which is a more encompassing term referring not only to PA but also to awareness of larger spoken units such as syllables and rhyming words. Although PA studies often begin by teaching children to analyze larger units, we did not include such studies in our data set unless the instruction progressed to the phonemic level.

Researchers have used the following tasks to assess children's PA and to improve their PA through instruction and practice:

1. Phoneme *isolation*, which requires recognizing individual sounds in words; for example, "Tell me the first sound in *paste*." (/p/)
2. Phoneme *identity*, which requires recognizing the common sound in different words; for example, "Tell me the sound that is the same in *bike*, *boy*, and *bell*." (/b/)
3. Phoneme *categorization*, which requires recognizing the word with the odd sound in a sequence of three or four words; for example, "Which word does not belong? *bus*, *bun*, *rug*." (*rug*)
4. Phoneme *blending*, which requires listening to a sequence of separately spoken sounds and combining them to form a recognizable word; for example, "What word is /s/ /k/ /u/ /l/?" (*school*)
5. Phoneme *segmentation*, which requires breaking a word into its sounds by tapping out or counting the sounds or by pronouncing and positioning a marker for each sound; for example, "How many phonemes in *ship*?" (3: /s/ /i/ /p/)
6. Phoneme *deletion*, which requires recognizing what word remains when a specified phoneme is removed; for example, "What is *smile* without the /s/?" (*mile*)

In a few studies, instruction was focused on teaching children to manipulate onsets and rimes in words (Fox & Routh, 1984; Lovett, Barron, Forbes, Cuksts, & Steinbach, 1994; Treiman & Baron, 1983; Wilson & Frederickson, 1995). The onset is the single consonant or blend that precedes the vowel, and the rime is the vowel and following consonants; for example, j-ump, st-op, str-ong. We included these studies because students were essentially manipulating phonemes when the onset was a single phoneme.

Some forms of PA instruction would qualify as phonics instruction, which involves teaching students how to use grapheme-phoneme correspondences to decode or spell words. What distinguished PA studies from the general pool of phonics studies, however, is that instruction given to students in the treatment groups, but withheld from controls, was limited to grapheme-phoneme manipulation and did not go beyond this to include other activities such as reading decodable text or writing stories.

Contribution of PA in learning to read. PA is one of the best predictors of how well children will learn to read. Share et al. (1984) assessed kindergartners on many measures when they entered school, including phonemic segmentation, letter name knowledge, memory for sentences, vocabulary, father's occupational status, parental reports of reading to children, and TV watching. They examined which of these measures best predicted how well the children would be reading at the end of kindergarten and at the end of first grade. Results showed that phonemic awareness was the top predictor along with letter knowledge. PA correlated $r = 0.66$ with reading achievement in kindergarten and $r = 0.62$ in first grade.

PA is thought to contribute to helping children learn to read because the structure of the English writing system is alphabetic. Moreover, it is not easy to figure out the system. Words have prescribed spellings that consist of graphemes symbolizing phonemes in predictable ways. Being able to distinguish the separate phonemes in pronunciations of words so that they can be matched up to graphemes is difficult. There are no breaks in speech signaling where one phoneme ends and the next one begins. Rather, phonemes are folded into one another and are coarticulated. Discovering phonemic units is helped greatly by explicit instruction in how the system works. This is underscored by research revealing that people who have not learned to read and write have great trouble performing phonemic awareness tasks (Morais, Bertelson, Cary, & Alegria, 1987). Likewise people who have learned to read in a script that is not graphophonemic, such as Chinese, have difficulty segmenting speech into phonemes (Mann, 1987; Read, Zhang, Nie, & Ding, 1987).

PA is thought to contribute to children's ability to read words in various ways (Ehri, 1991, 1994). Decoding words requires blending skill to transform graphemes into recognizable words. Reading words by analogy (e.g., reading *brick* by knowing how to read *kick*) requires onset-rime segmentation and blending skill. Reading words from memory by sight requires phoneme segmentation skill. To store individual sight words in memory, children need to match up graphemes to phonemes in the word and retain these connections in memory (Ehri, 1980, 1992; Ehri & Wilce, 1987a; Rack, Hulme, Snowling, & Wightman, 1994; Reitsma, 1983).

The impact of PA instruction on various types of word reading was examined in many of the studies. The simplest word reading task given to preschoolers required them to look at a word (*sat*) and decide whether it said "sat" or "mat" (Byrne & Fielding-Barnsley, 1991). Studies with older children assessed their ability to read a list of words or their ability to remember how to read words after practicing them. Pseudoword reading tasks assessed children's ability to read novel words such as *feem*, *bote*, and *cliss*.

PA is thought to help children write words in various ways. They may invent letter-sound spellings or retrieve correct spellings from memory. Both processes require phonemic segmentation skill, enabling spellers to match up sounds to letters in words (Griffith, 1991). The impact of PA instruction on spelling skill was assessed in several studies.

Some of the studies measured reading comprehension. To comprehend a text, readers must be able to read most of the words. However, other capabilities influence reading comprehension as well, such as readers' vocabulary, their world knowledge, and their memory for the text. We expected PA instruction to benefit children's reading comprehension because it depends on effective word reading. However, we did not expect the effect to be as strong on comprehension as on word reading because the influence is indirect.

Replication and extension of a previous meta-analysis. The present study attempted to replicate and extend findings of the meta-analysis of phonemic awareness instruction studies published by Bus and van IJzendoorn (1999), referred to here as the B&I study. Some of the same moderator variables were examined, including grade, reading ability, teaching with or without letters, instructor, size of the group taught, and whether the control group was treated or untreated. The present study expanded the database that tested the effects of PA instruction on reading outcomes from 34 studies to 52 studies. In the B&I analysis, when studies included more than one PA treatment or more than one control group, these groups were combined into single treatment-control comparisons. In contrast, we kept these comparisons separate in our analysis. As a result, the number of treatment-control comparisons increased from 34 in the B&I study to 96 in our study. Having more comparisons made it possible to expand the examination of moderator variables. Also spelling outcomes were evaluated in more studies. Having Troia's (1999) critique of PA studies allowed us to examine, more thoroughly than B&I could, the relationship between methodological design features and the size of effects detected in the experiments.

Methodological rigor of PA studies. Many correlational studies have reported strong relationships between phonemic awareness and learning to read (for reviews see Blachman, 2000; Ehri, 1979; Snow, Burns, & Griffin, 1998; Stahl & Murray, 1994; Wagner & Torgesen, 1987). However, correlational findings are insufficient to show that PA was the underlying *cause* enabling some students to read better than others. This is because correlational findings do not rule out other factors that might also explain the relationship. To show that PA operates as a direct cause in helping children learn to read, evidence from experimental studies with treatment and control groups is needed.

Well-designed experiments yielding positive outcomes provide the strongest evidence that PA caused the improvement in reading. Although all of the studies in our database consisted of experiments, some were better designed than others. Studies varied in whether they used treated or untreated control groups. The use of untreated controls receiving no special attention from researchers runs the risk of Hawthorne effects as an explanation for differences favoring the treatment group. Studies varied in whether students were randomly assigned to treatment and control groups, or whether a quasi-experimental design was used in which existing groups were assigned to conditions, or whether students were matched and assigned to conditions. Although random assignment is preferable, researchers may be limited to a quasi-experimental design when classrooms in schools are studied. We examined whether positive effects of PA instruction emerged primarily from the weaker designs or whether effects were strongest in the best designed experiments.

Surprisingly, B&I (1999) found larger effect sizes for comparisons that contrasted PA-instructed groups to *treated* control groups than to untreated control groups. This finding is the opposite of what would be expected if Hawthorne effects were inflating effect sizes. Whether this finding replicates in our database became a question of special interest.

In a recent critique of PA instruction studies, Troia (1999) identified several design flaws and applied these criteria to rate the studies for their methodological rigor. He considered threats to internal validity (i.e., attribution of cause to the treatment) as well as to external validity (i.e., generalizability of findings). To evaluate whether these flaws might be associated with outcomes, we examined the relationship between Troia's ratings and the effect sizes in the studies. Our purpose was to determine whether claims about PA instructional effects are supported mainly by poorly-designed or well-designed studies.

Other issues. One issue of interest was whether PA instruction might be more effective for some age/grade or reader groups than for others. Preschoolers use language as a tool for communication, so their focus is upon the meaning of speech, not upon its phonological structure. As a result they have little PA and hence stand to gain much from PA instruction. In contrast, beginning readers have acquired at least some awareness of phonemes, even without PA instruction, because making progress in reading requires grapheme-phoneme knowledge. PA instruction may still contribute to their growth in literacy, but its impact may be less than in the earliest grades. As readers and writers advance beyond first grade, the need to acquire additional PA may become less important than the need to learn about spelling patterns in words, so instruction focused on phonemes may yield diminishing returns. This suggests that the impact of PA instruction may be greatest in preschool and kindergarten, and may become smaller beyond first grade. In the B&I (1999) meta-analysis, although all groups profited from PA instruction, preschoolers benefited more than kindergartners or primary school students.

PA instruction may contribute less to older, normally developing readers, but it may make a big difference for older children who have failed to make normal progress in learning to read. Research has shown that disabled readers have poor phonemic awareness, even below that of nondisabled students reading at the same grade-equivalent level (Bradley & Bryant, 1983; Bruck, 1992; Fawcett & Nicholson, 1995). In addition, disabled readers have special difficulty learning to spell (Bruck, 1993). We might expect PA instruction to help in remediating the reading and spelling difficulties of these readers.

Instruction in PA may be conducted with or without alphabetic letters. In some studies, children were taught to manipulate phonemes in words by using letters as markers for the sounds whereas in other studies children were taught to work with spoken units only. Sounds are ephemeral, short-lived, and hard to grasp, whereas letters provide concrete, visible symbols for phonemes. Thus, we might expect children to have an easier time acquiring PA when they are given letters to manipulate. Also, because letters bring children closer to the task of applying PA in reading and spelling, we would expect transfer to be greater when PA is taught with letters. In the B&I (1999) study, PA instruction with letters produced larger effects on PA and reading than instruction without letters.

English includes many words that exhibit grapheme-phoneme regularity, but it also includes words that derive their regularity from larger, word-based spelling patterns or that contain exceptions to graphophonemic regularities. This contrasts with other written languages whose spellings are regular principally at the grapheme-phoneme level. The studies in our database provided PA instruction not only in English but also in other languages, namely, Norwegian, Finnish, Swedish, Danish, Spanish, Hebrew, Dutch, and German. In most of these languages, the grapheme-phoneme connections are more transparent than they are in English. Of interest was whether PA instruction might contribute more to reading acquisition in English because children need more help in figuring out a nontransparent graphophonemic system.

Because classroom teachers are the purveyors of reading instruction for most children, it is important to verify that they can teach PA effectively. Some of the studies used classroom teachers who taught PA to their students, thus allowing us to assess their effectiveness.

There is substantial evidence that 1:1 tutoring is the most effective form of instruction (Bloom, 1984; Cohen, Kulik, & Kulik, 1982; Glass, Cahen, Smith, & Filby, 1982; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; Wasik & Slavin, 1993). It allows instructors to tailor lessons to students' needs and difficulties. In the database, PA instruction was delivered not only to individual children but also to small groups and to classrooms of students. We expected tutoring to produce stronger effects than the other two types of delivery. However, B&I (1999) obtained only partial support for this. They found that individualized instruction was less effective than small-group instruction for teaching PA, but was more effective for promoting transfer to reading. Replication of this effect with our larger database was considered important.

Method

Database. We conducted an electronic search of two databases, ERIC and PsychInfo. Six terms involving phonemic awareness (PA) were crossed with 15 terms related to reading performance. The PA terms were *phonemic awareness, phonological awareness, spelling, blending, learning to spell,* and *invented spelling.* The reading terms were *reading, reading ability, reading achievement, reading comprehension, reading development, reading disabilities, reading skills, remedial reading, beginning reading, beginning reading instruction, reading acquisition, word identification, word reading, oral reading,* and *miscues.* Only articles appearing in journals written in English were considered. The search yielded 637 articles through ERIC and 1,325 articles through PsychInfo. In addition, we hand-searched references in the studies screened and in several review papers (Apthorp, 1998; Blachman, 2000; Bus & van IJzendoorn, 1999; Stahl & Murray, 1994; Troia, 1999; Wagner, 1988).

To qualify for our analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group.
2. Studies had to appear in a refereed journal.
3. Studies had to test the hypothesis that instruction in phonemic awareness improves reading performance over alternative forms of instruction or no instruction.
4. Studies had to provide instruction in phonemic awareness that was not confounded with other instructional methods or activities.
5. Studies had to report statistics permitting the calculation or estimation of effect sizes.

The requirement that studies come from refereed journals was adopted by the National Reading Panel and applied to all the meta-analyses. Our rationale was that limiting studies to those passing the test of peer review serves to minimize the risk of admitting studies of poor quality.

The final set of studies meeting our criteria numbered 52. From these, 96 cases comparing individual treatment and control groups were derived. Because some of the studies included multiple treatment or control groups, our cases included comparisons utilizing the same group more than once. Seven treatment groups were included twice because they were compared to two different control groups. Sixteen control groups were included twice because they were compared to two different treatment groups. One control group appeared three times because it was compared to three treatment groups. Although this meant that effect sizes were not completely independent across cases, we preferred this alternative to that of combining treatment and control groups within studies because we did not want to obscure important moderator variables of interest.

The following studies included treatment or control groups that were not deemed appropriate for our analysis, so these comparisons were not included: a treatment group given decoding instruction and word reading (Barker & Torgesen, 1995); a treatment group given a reading and writing program (Brennan & Ireson, 1997); a treatment group taught to manipulate syllables rather than phonemes (Sanchez & Rueda, 1991); a treatment group taught semantic categorization with written words (Defior & Tudela, 1994); treatment groups in which the instructors failed to spend the time prescribed for training (Olofsson & Lundberg, 1983); treatment groups in which children not only analyzed phonemes but also read words in sentences and stories, unlike children in the control groups who only listened to stories or remained in their classrooms (Solity, 1996; Weiner, 1994); a control group lacking not only PA instruction but also the Reading Recovery instruction given the treatment group (Iversen & Tunmer, 1993); a control group that did not control for all of the non-PA elements of instruction (Lovett et al., 1994; Vellutino & Scanlon, 1987).

The studies in our database were coded for several characteristics. These are listed in Table 1. Three outcomes were of primary interest: phonemic awareness, reading, and spelling. Some studies measured one or another of the three outcomes in multiple ways and reported means and standard deviations on each measure; for example, measuring children's ability to segment words into phonemes, to delete phonemes from words, and to blend phonemes into words. Because we wanted one overall effect size for each outcome from each study, we combined measures of the same outcome within each study by first calculating effect sizes on each measure and then averaging effect sizes across measures to create one effect size for the outcome of interest. The combined measure for phonemic awareness was limited to tasks assessing phoneme manipulation, not larger unit manipulation. The kinds of measures that were combined into the reading outcome included word reading, pseudoword reading, reading comprehension, oral text reading, reading speed, time to reach a criterion of learning, and miscues. The spelling composite included measures of invented spellings as well as correct spellings of words and pseudowords. In this way, each treatment-control group comparison had the potential for contributing one effect size to the pool of effect sizes for phonemic awareness, for reading, and for spelling outcomes.

In addition, we were interested in the effect of PA instruction on specific measures of PA (i.e., segmentation, blending, deletion, and other), on specific measures of reading (i.e., words, nonwords, comprehension), and on outcomes measured immediately after testing as well as after a delay. When studies administered delayed posttests at more than one point in time, we calculated effect sizes for the first two test points. Separate effect sizes on these measures were included for each study in the database. Also we compared effect sizes on experimenter-devised and standardized tests of reading and spelling. Standardized tests are valued because they give a more general picture of transfer effects. However, when standardized tests are designed to distinguish individuals across many ages or grades, they may be less sensitive to differences within a narrow range of performance.

Because several studies examined whether PA instruction impacted students' performance in math, we included this as well. It was expected that scores in math would not show gains, hence ruling out Hawthorne effects.

There were three reader groups. One group consisted of children who were not distinguished as having any reading problems, referred to as normally developing readers. One group consisted of children below second grade who were at risk for developing reading difficulties. They were given this label by authors of the studies. Being at risk was indicated by low PA or low reading in 83% of the cases and low socioeconomic status in 27% of the cases. Other at-risk indicators were developmental or language delays or cognitive disabilities.

The third reader group consisted of reading disabled (RD) students who had already developed reading problems. All but three cases involved children between second and sixth grades. The three cases were first graders who qualified for Reading Recovery programs (Hatcher, Hulme, & Ellis, 1994; Iversen & Tunmer, 1993). In most studies, being reading disabled meant reading below grade level despite at least average cognitive ability. In one study, the school's definition of learning disabled was used (Williams, 1980). In one study, students had not only RD but also neurological impairment and language-learning problems (Lovett et al., 1994).

PA programs varied in whether they focused on specific PA manipulations. Single-focus studies taught blending, categorization, identity, segmentation, or onset-rime. Double-focus studies involved combinations of blending, segmenting, deletion, or categorization. Global treatments taught three or more PA skills. Programs that taught onset-rime manipulation were coded as having a single focus even though the instruction might have taught children to blend or segment the onsets and rimes (e.g., Fox & Routh, 1976). Instruction varied in whether children were taught to manipulate phonemes using letters or whether attention was limited to phonemes in speech. Instruction that had children manipulate blank markers was coded as a nonletter treatment.

The instructional delivery unit varied across studies. Students were tutored individually or taught in small groups or whole classrooms. The size of the small groups varied from two to seven students. The identity of instructors varied across studies and included classroom teachers, researchers or their assistants, or computers. Credentialed teachers who conducted the instruction but were not the students' classroom teacher were coded as researchers. We coded studies to reflect whether fidelity to treatment was checked, that is, whether researchers observed instructors to assess their adherence to treatment procedures.

Some features of the experimental methodology were coded. One feature was group assignment. In some studies, children were randomly assigned to conditions. In some studies, children were members of intact groups that were assigned to conditions, referred to as nonequivalent groups. When two classrooms were assigned randomly, one to the treatment and one to the control condition, they were categorized as nonequivalent groups. When several classrooms were assigned randomly to treatment and control conditions, they were categorized as random assignment. The third way of assigning children to conditions involved matching children on the basis of similar test scores. Although members of a match

Table 1 Dependent and moderator variables included in the meta-analyses

Outcome measures

1. Composite measures: phonemic awareness; reading; spelling
2. Measures of phonemic awareness: segmentation; blending; deletion; other
3. Measures of reading:
    - Standardized versus experimenter-devised tests of word reading
    - Standardized versus experimenter-devised tests of nonword reading
    - Reading comprehension
4. Measures of spelling: Standardized versus experimenter-devised tests of spelling
5. Measure of math achievement
6. Test points:
    - Immediately after instruction
    - First follow-up test (delay of 2 to 15 months)
    - Second follow-up test (delay of 7 to 36 months)

Characteristics of participants

1. Reader level: at-risk readers; disabled readers; normally progressing readers
2. Grade level: preschool; kindergarten; first grade; second through sixth grades
3. Socioeconomic status: low-SES; middle-to-high-SES
4. Language: English; other (Danish, Dutch, Finnish, German, Hebrew, Norwegian, Spanish, Swedish)

Properties of phonemic awareness instruction

1. PA skills taught:
    - a. single skill; two skills; three or more skills
    - b. segmenting and blending versus three or more skills
2. Use of letters: phonemes and letters manipulated versus only phonemes manipulated
3. Delivery unit: individuals; small groups (two to seven students); classrooms
4. Identity of instructor: classroom teachers; computers; researchers/others
5. Length of instruction: ranged from 1 hour to 75 hours

Features of the design

1. Group assignment: random, matched; nonequivalent
2. Fidelity of instructors: checked versus not checked or not reported
3. Control group: alternative treatment; no treatment
4. Size of the sample: ranged from 9 to 383 students
5. Internal validity (from Troia, 1999):
    - Percentage of criteria met
    - Number of critical flaws
6. External validity (from Troia, 1999):
    - Percentage of criteria met
    - Number of critical flaws
7. Methodological rigor (from Troia, 1999):
    - Overall ranking

Characteristics of the study

Year of publication (1976 to 2000)

are typically assigned randomly to conditions, in some studies this step was not stated explicitly.

To evaluate the relationship between the methodological quality of studies and the effect sizes found, we adopted the five methodological criteria applied by Troia (1999) in his critique of the internal and external validity of PA studies. Internal validity refers to whether the treatment rather than some other variable caused the outcome observed. External validity refers to the generalizability of the findings, that is, whether or not the results of a study can be applied to other persons in other settings at other times. To evaluate the internal and external validity of studies, Troia used five summary measures: percentage of internal validity criteria met by the studies, number of critical flaws challenging a study's internal validity (e.g., no random assignment, no alternative treatment given to the control group, no assessment of the instructor's fidelity to treatment), percentage of external validity criteria met, number of critical flaws challenging a study's external validity (e.g., insufficient information about the sample of participants or about how disability was defined and assessed). Troia evaluated 28 of the studies included in our database. We applied his ratings and rankings to the 56 comparisons derived from these studies. We did this without checking Troia's evaluations for accuracy, so any incorrect codings of the studies are the result of Troia's procedures, not ours.

Four people coded the studies and entered values into the SPSS (Statistical Package for the Social Sciences) database. The reliability of moderator-variable codes was checked by comparing codes in the database to codes generated by one of the coders who recoded 14 of the articles (15% of the cases). The percentage of agreement of the codes was 94%. All of the means, standard deviations, and sample sizes entered into the database were verified at least twice for accuracy.

Meta-analysis. The primary statistic used in our analysis of outcomes 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 was the mean of the treatment group minus the mean of the control group divided by a pooled standard deviation. Use of the pooled *SD* was adopted by the National Reading Panel for use in all of its meta-analyses.

Most of our studies reported treatment and control group means and standard deviations, which were used to calculate effect sizes. However, there were 14 studies that lacked sufficient information. We employed DSTAT (Johnson, 1989) to estimate these effects, usually from *F*- or *t*- or *MSE* values, or we obtained the information from authors.

Some studies included multiple measures of PA, reading, or spelling. To ensure that each treatment-control comparison contributed only one effect size to each of the three outcomes, we calculated a raw effect size (g) for each measure when several were taken, and then we averaged the effect sizes across these measures to create one composite effect size for each outcome.

The analysis of effect sizes across comparisons was conducted by giving more weight to effect sizes that were based on larger samples of participants. However, a few studies administered instruction to groups of students and hence used groups rather than individual students as the unit of analysis in their statistics (Byrne & Fielding-Barnsley, 1991; Castle, Riach, & Nicholson, 1994; O'Connor, Jenkins, & Slocum, 1995; Torgesen, Morgan, & Davis, 1992; Williams, 1980, Experiment 2). Using the number of groups in the weighting procedure for these comparisons had the effect of underrepresenting their effect sizes relative to effect sizes of the other comparisons where the number of participants was used. To address this problem, we used the unit of analysis to convert raw effect sizes (g) to corrected effect sizes (d) for each treatment-control comparison. Then when we calculated mean effect sizes across comparisons, we weighted the individual effect sizes (d) by the number of students in the sample, not by the unit of analysis, thus ensuring that no cases were underrepresented.

We employed the DSTAT statistical package (Johnson, 1989) to determine effect sizes and to test the influence of moderator variables on effect sizes. Each moderator variable had at least two levels. We tested whether the mean weighted effect size (d) at each level was statistically greater than zero at $p < .05$, whether the individual effect sizes at each level were statistically homogeneous ($p < .05$), and whether effect sizes differed statistically at different levels of the moderator variables ($p < .05$). We received assistance from two consultants in the conduct of our meta-analysis, Blair Johnson and Harris Cooper, both of whom have authored papers or books on the subject.

Results

Were effect sizes greater than zero?

The effect size statistic measures how much the mean of the PA-instructed group exceeded the mean of the control group in standard deviation units. An effect size of 1 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of instruction. An effect size of 0 indicates that treatment and control group means were identical, showing that instruction had no effect. To judge the strength of an effect size, values suggested by Cohen (1988) are commonly used. An effect size of $d = 0.20$ is considered small, an effect size of $d = 0.50$ moderate, and an effect size of $d = 0.80$ or above large. Translated into more familiar terms, $d = 0.20$ 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.

Mean effect sizes (d) associated with levels of the moderator variables are reported in Table 2 for phonemic awareness outcomes, in Table 3 for reading outcomes, and in Table 4 for spelling outcomes. Values of d that were statistically greater than zero are marked with an asterisk. Inspection of values in Tables 2 and 3 reveals that all of the effect sizes involving PA and reading outcomes were statistically greater than zero ($p < .05$). This indicates that instruction was uniformly effective in teaching PA and in facilitating transfer to reading across all levels of the moderator variables that were considered.

The overall effect size of PA instruction on the acquisition of PA was large, $d = 0.86$, based on 72 comparisons. This is somewhat less than the effect size detected by B&I (1999) who reported $d = 1.04$ based on 36 comparisons. The overall effect size on reading was moderate, $d = 0.53$, based on 90 comparisons. This is slightly greater than the effect size reported by B&I who found $d = 0.44$, based on 34 cases. Although not identical, the values are comparable, indicating that our larger database did not alter earlier overall findings.

Inspection of spelling outcomes in Table 4 reveals that all but three effect sizes were statistically greater than zero. This indicates that, across most levels of the moderator variables, PA instruction transferred and improved spelling skills more than alternative forms of instruction or no instruction.

Some of the studies evaluated effects on an outcome not expected to be affected by PA instruction, performance in math. Math posttests were administered immediately after instruction in 12 comparisons and after some delay in three comparisons. Results in Table 3 show that the effect size was statistically nonsignificant and close to zero ($d = 0.03$), indicating that the effects of PA instruction were limited to literacy outcomes. These findings argue against the operation of any halo or Hawthorne effect explaining the positive effect sizes.

In sum, these findings lead us to conclude with much confidence that phonemic awareness instruction is more effective than alternative forms of instruction or no instruction in helping children acquire phonemic awareness and in facilitating transfer of PA skills to reading and spelling.

Were effect sizes homogeneous?

A homogeneity analysis calculates how probable it is that the variance exhibited among the effect sizes would be observed if only sampling error was making them different (Cooper, 1998). If found homogeneous, the mean effect size can be interpreted as representative of that set. When effect sizes are not homogeneous, the next step is to examine whether moderator variables create homogeneity, indicating their power to explain the variance.

Tables 2, 3, and 4 report results of the test of homogeneity ($p < .05$) and 95% confidence intervals to reveal how variable effect sizes were. At the tops of the tables, it is apparent that on the immediate outcome measures of PA, reading and spelling, effect sizes were not homogeneous. Effect sizes involving follow-up measures of PA and spelling outcomes were homogeneous, but follow-up reading effect sizes were not. Thus, there is reason to examine moderator variables that may explain effects on immediate outcomes of all three variables and on follow-up tests of reading.

Did moderator variables influence effect sizes?

Effects of several moderator variables were examined (see list in Table 1). Some cautions should be applied in interpreting the effects of moderators. When comparing the size of effects associated with levels of moderators, one must remember that PA instruction was effective for almost all the levels examined, so one should not mistakenly equate smaller effects with statistically nonsignificant effects.

Also when effect sizes are larger for some levels of a moderator variable than for others, it is not clear that the moderator caused PA instruction to vary in its impact rather than a third hidden factor confounded with the moderator. Likewise, if a moderator exerts no differential effect, it may be that a third factor obscured the difference. This renders conclusions about the influence of moderators tentative and suggestive rather than definitive.

Another caution is that all 96 treatment-control comparisons in our database did not contribute to the calculation of every effect size. Rather the comparisons changed across moderator variables, either because studies did not report the information, or they did not assess that outcome at that test point. Thus, one cannot assume that effect sizes represent the whole database.

Outcome measures. From Table 2, it is apparent that PA instruction was effective in improving children's phonemic awareness. The effect size after instruction was large ($d = 0.86$), and this value did not differ statistically

Table 2 Phonemic awareness outcomes: Mean effect sizes (d) as a function of moderator variables and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$. Effect sizes were measured immediately after instruction unless labeled as follow-up

Moderator variables and levels	Number of cases	Mean d	Homogeneous	95% confidence interval	Contrasts[a]
Time of posttest					
Immediate	72	.86*	No	.79 to .92	*ns*
Follow-up	14	.73*	Yes	.61 to .85	
Outcome measures of PA					
Segmentation	51	.87*	No	.79 to .94	S = D > B
Blending	33	.61*	No	.52 to .69	S > O
Deletion	25	.82*	No	.73 to .91	B = O
Other	37	.72*	No	.64 to .81	D = O
Characteristics of participants					
Reading level					
At risk	15	.95*	No	.76 to 1.14	A = N > D
Disabled	15	.62*	No	.48 to .75	
Normal progress	42	.93*	No	.85 to 1.01	
Grade					
Preschool	2	2.37*	No	1.93 to 2.81	P > K > 1 = 2
Kindergarten	39	.95*	No	.87 to 1.04	
First	15	.48*	Yes	.31 to .64	
Second–sixth	16	.70*	Yes	.56 to .83	
Socioeconomic status					
Low	12	1.07*	No	.93 to 1.20	*ns*
Middle and high	17	1.02*	No	.87 to 1.18	
Language of instruction					
English	61	.99*	No	.90 to 1.07	E > O
Other	11	.65*	Yes	.55 to .76	
Characteristics of PA instruction					
Skills taught					
One skill	18	1.16*	No	.96 to 1.36	1 = 2 > 3
Two skills	24	1.03*	No	.92 to 1.14	
Three or more skills	30	.70*	No	.61 to .78	
Blend and segment only	18	.81*	No	.67 to .95	*ns*
Three or more skills	30	.70*	No	.61 to .78	
Use of letters					
Letters manipulated	39	.89*	No	.80 to .98	*ns*
Letters not manipulated	33	.82*	No	.73 to .91	
Delivery unit					
Individual child	24	.60*	Yes	.47 to .72	S > I = C
Small groups	35	1.38*	No	1.26 to 1.50	
Classrooms	13	.67*	No	.57 to .76	
Length of instruction					
1 to 4.5 hours	15	.61*	Yes	.41 to .81	5 = 10 > 1 = 20
5 to 9.3 hours	24	1.37*	No	1.23 to 1.51	
10 to 18 hours	9	1.14*	No	.97 to 1.32	
20 to 75 hours	22	.65*	No	.56 to .74	
Characteristics of instructors					
Classroom teachers	19	.78*	No	.70 to .87	RO > CT
Researchers and others	53	.94*	No	.84 to 1.03	
Computers	8	.66*	Yes	.52 to .85	O > C
Others	64	.89*	No	.82 to .96	

(continued)

Table 2 Phonemic awareness outcomes (continued)

Moderator variables and levels	Number of cases	Mean *d*	Homogeneous	95% confidence interval	Contrasts[a]
Characteristics of design					
Random assignment	33	.87*	No	.77 to .97	*ns*
Matched	18	.92*	No	.75 to 1.09	
Nonequivalent	21	.83*	No	.73 to .92	
Fidelity checked	29	.66*	No	.56 to .75	N > F
Fidelity not checked	43	1.02*	No	.93 to 1.11	
Treated controls	38	.89*	No	.79 to .99	*ns*
Untreated controls	34	.83*	No	.75 to .92	
Size of sample					
9 to 22 students	15	1.37*	No	1.09 to 1.66	9 = 31 > 24 = 56
24 to 30 students	22	.70*	No	.53 to .87	
31 to 53 students	13	1.10*	No	.90 to 1.30	
56 to 383 students	22	.82*	No	.74 to .89	
Characteristics of study					
Year of publication					
1976–1985	10	.73*	Yes	.53 to .94	1991 > others
1986–1990	16	.72*	No	.59 to .85	1976 = 1986 = 1996
1991–1995	31	1.18*	No	1.07 to 1.30	
1996–2000	15	.70*	No	.59 to .81	

Note. *d* = mean effect size; Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous; PA = phonemic awareness; *ns* = results were not statistically significant.
[a] Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters/numbers of levels.
* Indicates that effect size was statistically greater than zero at $p < .05$.

from the effect size at the follow-up test ($d = 0.73$), indicating no significant long-term decline. Thus, PA instruction taught phonemic awareness very effectively, and students retained their skill after instruction ended. Comparison of specific PA skills acquired during instruction indicated that effects were statistically larger for segmentation and deletion outcomes than for blending.

The strong gains in PA transferred to reading, and effects persisted through the second follow-up test. As evident in Table 3, reading outcome effect sizes were moderate, and the effect size at the end of instruction ($d = 0.53$) did not differ statistically from that at the first follow-up test ($d = 0.45$). A statistically significant effect size was still present but statistically smaller at the second follow-up test ($d = 0.23$). The type of test made a difference. The effect size was statistically larger on experimenter tests than on standardized tests.

Some studies measured reading performance with pseudowords to assess children's ability to decode unfamiliar words. From Table 3, it is apparent that PA instruction benefited decoding skill ($d = 0.56$). Effects were statistically greater than zero, were moderate, and did not differ statistically on experimenter-devised tests and standardized tests.

The effect of PA instruction on reading comprehension was assessed in 20 comparisons. From Table 3, it is apparent that the effect size was statistically greater than zero ($d = 0.34$), indicating that PA instruction exerted a small-to-moderate impact on readers' ability to comprehend text.

PA instruction also transferred to spelling (see Table 4). The effect size following instruction ($d = 0.59$) was moderate and statistically greater than the effect sizes at the two delayed posttests ($d = 0.37$ and 0.20), both of which were statistically greater than zero and did not differ statistically. The effect size was statistically larger on experimenter tests than on standardized tests of spelling.

Characteristics of students. Three types of readers were distinguished: at risk, disabled (RD), and normally progressing readers. Younger students at risk for developing reading problems and older disabled readers have been found to exhibit excessive difficulty manipulating phonemes in words (Bradley & Bryant, 1983; Juel, 1988; Juel, Griffith, & Gough, 1986), so we were especially interested in the impact of PA instruction on these readers. A comparison of PA outcomes across the three reader groups revealed that although effect sizes were moderate to large in all cases, they were statistically larger for at-risk readers ($d = 0.95$) and normally progressing readers ($d = 0.93$) than for disabled readers ($d = 0.62$). Children at risk gained as much PA, according to statistical tests, as normally developing readers, indicating that at-risk read-

Table 3 Reading outcomes: Mean effect sizes (*d*) as a function of moderator variables and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$. Effect sizes were measured immediately after instruction unless labeled as follow-up

Moderator variables and levels	Number of cases	Mean *d*	Homogeneous	95% confidence interval	Contrasts[a]
Characteristics of outcome measures					
Time of posttest					
Immediate	90	.53*	No	.47 to .58	Im = 1 > 2
First follow-up	35	.45*	No	.36 to .54	
Second follow-up	8	.23*	No	.11 to .34	
Type of word test					
Experimenter	58	.61*	No	.54 to .69	E > S
Standardized	37	.32*	No	.23 to .42	
Type of pseudoword test					
Experimenter	47	.56*	No	.48 to .64	*ns*
Standardized	8	.49*	Yes	.29 to .69	
Reading comprehension	20	.34*	No	.21 to .46	
Math achievement	15	.03*ns*	No	-.11 to .16	
Characteristics of participants					
Reading level					
Immediate posttest					
At risk	27	.86*	No	.72 to 1.00	A > D = N
Disabled	17	.45*	Yes	.32 to .57	
Normal progress	46	.47*	No	.39 to .54	
Follow-up posttest					
At risk	15	1.33*	No	1.10 to 1.56	A > D = N
Disabled	8	.28*	Yes	.10 to .46	
Normal progress	12	.30*	Yes	.19 to .42	
Grade					
Preschool	7	1.25*	No	1.01 to 1.50	P > K = 1 = 2
Kindergarten	40	.48*	No	.40 to .56	
First	25	.49*	Yes	.36 to .62	
Second–sixth	18	.49*	Yes	.35 to .62	
Socioeconomic status					
Low	11	.45*	No	.33 to .58	MH > L
Middle and high	29	.84*	No	.72 to .96	
Language of instruction					
Immediate posttest					
English	72	.63*	No	.55 to .70	E > O
Other	18	.36*	No	.27 to .46	
Follow-up posttest					
English	17	.42*	Yes	.28 to .56	*ns*
Other	18	.47*	No	.35 to .59	
Characteristics of PA instruction					
Skills taught					
Immediate posttest					
One skill	32	.71*	No	.58 to .84	1 = 2 > 3
Two skills	29	.79*	No	.69 to .89	
Three or more skills	29	.27*	Yes	.19 to .35	
Follow-up posttest					
One skill	11	.55*	Yes	.37 to .73	2 > 1 > 3
Two skills	9	1.28*	No	.56 to .89	
Three or more skills	15	.23*	Yes	.11 to .37	
Blend and segment only	19	.67*	No	.54 to .81	BS > 3
Three or more skills	29	.27*	Yes	.19 to .35	
Use of letters					
Immediate posttest					
Letters manipulated	48	.67*	No	.59 to .75	L > NoL
Letters not manipulated	42	.38*	No	.30 to .46	
Follow-up posttest					
Letters manipulated	16	.59*	No	.45 to .74	L > NoL
Letters not manipulated	19	.36*	No	.25 to .47	

(continued)

Table 3 Reading outcomes (continued)

Moderator variables and levels	Number of cases	Mean d	Homogeneous	95% confidence interval	Contrasts
Delivery unit					
Immediate posttest					
Individual child	32	.45*	Yes	.34 to .57	S > I = C
Small groups	42	.81*	No	.71 to .92	
Classrooms	16	.35*	No	.26 to .44	
Follow-up posttest					
Individual child	7	.33*	Yes	.11 to .55	S > I = C
Small groups	18	.83*	No	.66 to 1.00	
Classrooms	10	.30*	Yes	.18 to .42	
Length of instruction					
1 to 4.5 hours	17	.61*	Yes	.42 to .79	1 = 5 = 10
5 to 9.3 hours	23	.76*	No	.62 to .89	20 < others
10 to 18 hours	19	.86*	No	.72 to 1.00	
20 to 75 hours	25	.31*	No	.22 to .39	
Characteristics of instructors					
Immediate posttest					
Classroom teachers	22	.41*	No	.33 to .49	RO > CT
Researchers and others	68	.64*	No	.56 to .73	
Follow-up posttest					
Classroom teachers	12	.32*	Yes	.20 to .43	RO > CT
Researchers and others	23	.63*	No	.49 to .77	
Computers	8	.33*	Yes	.16 to .49	O > C
Others	82	.55*	No	.49 to .61	
Characteristics of design					
Random assignment	46	.63*	No	.54 to .72	R > N
Matched	22	.57*	Yes	.43 to .72	M = all
Nonequivalent	20	.40*	No	.31 to .49	
Fidelity checked	31	.43*	No	.34 to .53	N > F
Fidelity not checked	59	.59*	No	.51 to .66	
Control group					
Immediate posttest					
Treated controls	54	.65*	No	.56 to .73	T > U
Untreated controls	36	.41*	No	.33 to .49	
Follow-up posttest					
Treated controls	20	.62*	No	.48 to .75	T > U
Untreated controls	15	.32*	Yes	.20 to .44	
Size of sample					
9 to 22 students	24	.72*	No	.51 to .92	9 = 31 > 56
24 to 30 students	22	.54*	Yes	.37 to .70	24 = 9,56
31 to 53 students	22	.91*	No	.76 to 1.05	31 > 24
56 to 383 students	22	.40*	No	.33 to .48	
Characteristics of Study					
Year of publication					
1976–1985	20	.77*	No	.62 to .93	1976 = 1991 >
1986–1990	16	.36*	Yes	.24 to .49	1986 = 1996
1991–1995	41	.77*	No	.67 to .87	
1996–2000	13	.21*	Yes	.11 to .32	

Note. d = mean effect size; Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous; PA = phonemic awareness; *ns* = results were not statistically significant.

[a] Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters/numbers of levels.

* Indicates that effect size was statistically greater than zero at $p < .05$.

ers were not any less responsive to instruction in acquiring PA. Disabled readers may have exhibited smaller effect sizes because they were older and relatively more advanced in PA skills with less room for gains than the younger readers, and also because they were taught more advanced forms of PA that may be harder to acquire.

Transfer of PA instruction to reading was also influenced by reader ability. Table 3 reveals that at-risk children showed statistically larger transfer effects on reading (d = 0.86) than normal and RD students whose effect sizes did not differ statistically (d = 0.47 for normals and d = 0.45 for RD). Effect sizes on follow-up reading tests

Table 4 Spelling outcomes: mean effect sizes (*d*) as a function of moderator variables and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$. Effect sizes were measured immediately after instruction unless labeled as follow-up

Moderator variables and levels	Number of cases	Mean *d*	Homogeneous	95% confidence interval	Contrasts[d]
Characteristics of outcome measures					
Time of posttest					
Immediate	39	.59*	No	.49 to .68	Im > 1 = 2
First follow-up	17	.37*	Yes	.26 to .48	
Second follow-up	6	.20*	No	.08 to .32	
Type of spelling test					
Experimenter	24	.75*	No	.62 to .89	E > S
Standardized	20	.41*	No	.29 to .53	
Characteristics of participants					
Reading level					
At risk	13	.76*	No	.54 to .98	A = N > D
Disabled	11	.15*ns*	Yes	-.00 to .31	
Normal progress	15	.88*	No	.74 to 1.02	
Grade					
Preschool	0	—	—	—	
Kindergarten	15	.97*	No	.82 to 1.13	K > 1 > 2
First	16	.52*	No	.37 to .68	
Second–sixth	8	.14*ns*	Yes	-.04 to .33	
Socioeconomic status					
Low	6	.76*	Yes	.57 to .95	MH > L
Middle and high	9	1.17*	No	.88 to 1.47	
Language of instruction					
English	32	.60*	No	.49 to .70	*ns*
Other	7	.55*	Yes	.31 to .78	
Characteristics of PA instruction					
PA skills taught					
One skill	17	.74*	No	.56 to .92	1 = 2 > 3
Two skills	12	.87*	Yes	.71 to 1.03	
Three or more skills	10	.23*	No	.07 to .38	
Blend and segment only	7	.79*	Yes	.49 to 1.09	BS > 3
Three or more skills	10	.23*	No	.07 to .38	
Use of letters					
Letters manipulated	27	.61*	No	.50 to .72	L > NoL
Letters not manipulated	12	.34*	No	.25 to .42	
Delivery unit					
Individual child	14	.36*	No	.20 to .52	S > I
Small groups	20	.77*	No	.63 to .90	C = all
Classrooms	5	.56*	No	.33 to .78	
Length of instruction					
1 to 4.5 hours	0	—	—	—	
5 to 9.3 hours	8	1.13*	Yes	.86 to 1.39	5 = 10 > 20
10 to 18 hours	10	.87*	No	.69 to 1.05	
20 to 75 hours	18	.32*	No	.19 to .45	
Characteristics of instructors					
Classroom teachers	9	.74*	No	.58 to .90	CT > RO
Researchers and others	30	.51*	No	.39 to .62	
Computers	6	.09*ns*	Yes	-.10 to .28	O > C
Others	33	.74*	No	.63 to .85	

(continued)

Table 4 Spelling outcomes (continued)

Moderator variables and levels	Number of cases	Mean d	Homogeneous	95% confidence interval	Contrasts[a]
Characteristics of design					
Random assignment	17	.37*	No	.23 to .50	M = N > R
Matched	12	.73*	No	.52 to .93	
Nonequivalent	10	.86*	Yes	.69 to 1.04	
Fidelity checked	15	.44*	No	.30 to .59	N > F
Fidelity not checked	24	.69*	No	.57 to .81	
Treated controls	24	.43*	No	.30 to .55	U > T
Untreated controls	15	.82*	No	.67 to .96	
Size of sample					
9 to 22 students	15	.85*	Yes	.59 to 1.10	24 > all
24 to 30 students	3	1.68*	Yes	1.15 to 2.21	9 > 56
31 to 53 students	8	.75*	No	.51 to .98	31 = 9.56
56 to 383 students	13	.45*	No	.34 to .56	

Note. d = mean effect size; Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous; PA = phonemic awareness; *ns* = results were not statistically significant.

[a] Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters/numbers of levels.

* Indicates that effect size was statistically greater than zero at $p < .05$.

showed the same pattern except that the effect size for at-risk students grew even larger ($d = 1.33$) while the effect sizes for the other two groups grew smaller ($d = 0.30$ for normals and 0.28 for RD). These findings indicate that PA instruction gave at-risk students a bigger boost in reading than it gave normals or disabled readers.

Transfer of PA instruction to spelling was also influenced by reader ability. Effect sizes were large and did not differ statistically for at-risk ($d = 0.76$) and normal readers ($d = 0.88$), indicating that PA instruction strongly benefited spelling for these students. However, the effect size was small and not statistically different from zero for disabled readers ($d = 0.15$). Moreover, the set of effect sizes for disabled readers was statistically homogeneous, indicating that no further analysis of moderator variables was needed to conclude that PA instruction did not help RD students to spell.

The effects of PA instruction were examined at various grade levels: preschool, kindergarten, first grade, and second–sixth grades. It should be noted that 78% of the second–sixth grade comparisons (i.e., 14 out of 18) involved disabled readers, so findings apply mainly to RD students and not to second–sixth graders in general. In acquiring PA, preschoolers showed a very large effect size ($d = 2.37$), although only two comparisons contributed to this value, making it less reliable. The effect on PA outcomes in kindergarten ($d = 0.95$) was statistically larger than the effect in first grade ($d = 0.48$) and in second–sixth grades ($d = 0.70$). The latter two effect sizes did not differ statistically. These findings indicate that preschoolers and kindergartners gained the most PA, not surprisingly because they started out with the least PA.

Effect sizes for reading outcomes in Table 3 reveal that PA instruction transferred to reading to a similar extent for kindergartners, first graders, and second–sixth graders. Effect sizes were all statistically greater than zero and did not differ statistically (ds = 0.48 to 0.49). The effect size for preschoolers was much larger statistically ($d = 1.25$), based on 7 comparisons and tested with simplified word recognition tests.

Transfer of PA instruction to spelling was statistically greater among kindergartners ($d = 0.97$) than among first graders ($d = 0.52$). There was no transfer to spelling among the second–sixth graders for whom the effect size ($d = 0.14$) did not differ from zero statistically. Spelling was not measured in the preschool studies. The absence of an effect on spelling among the older children arose because all of the comparisons involved disabled readers who did not gain in spelling from PA instruction.

SES levels did not have an impact on the acquisition of PA, but they did influence reading and spelling outcomes. As evident in Table 2, both low-SES and mid–high-SES groups showed large effect sizes that did not differ significantly in learning PA. However, transfer to reading and spelling was statistically greater among mid–high than among low-SES students (see Tables 3 and 4). It might be noted that most studies of disabled readers did not report the students' SES, so effect sizes are based primarily on the SES of normally developing and at-risk children.

Studies examining PA instruction were conducted not only in English-speaking countries but also in countries speaking languages other than English. A comparison of effect sizes revealed that PA instruction exerted a statistically larger impact on the acquisition of PA by

English-speaking students (d = 0.99) than by the non-English-speaking students (d = 0.65). Transfer to reading was also statistically greater for English-speaking students (d = 0.63) than for others (d = 0.36) on the immediate test but not on the follow-up test. Effect sizes on spelling outcomes did not differ statistically in the two language groups. One possible reason is that 94% of the RD comparisons were in the English pool, possibly suppressing the English spelling effect size. When the effect size was recalculated with the RD comparisons removed, a difference emerged. As evident in Table 5, the English comparisons yielded a statistically larger effect size on spelling (d = 0.95) than the non-English comparisons (d = 0.51). One possible reason for the larger effect sizes in English may be that the English writing system is not as transparent in representing phonemes as it is in the majority of the other languages, so explicit PA instruction may make a bigger contribution to clarifying phoneme units and their linkage to graphemes in English.

Because results of the language moderator variable changed when RD comparisons were removed from the analysis, we examined whether results might change for other moderators having uneven distributions of disabled readers across their levels. Inspection of distributions revealed some potential cases. Disabled readers were older (mostly in Grades 2 through 6), they tended to receive PA

Table 5 Mean effect sizes (d) with reading disabled comparisons removed from the database and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$.

Moderator variables and levels	Number of cases	Mean d	Homogeneous	95% confidence interval	Contrasts[a]
Spelling outcomes					
Grade					
Preschool	0	—			
Kindergarten	15	.97*	No	.82 to 1.13	K > 1
First	13	.66*	No	.48 to .85	
Second–sixth	0	—			
Language of instruction					
English	22	.95*	No	.82 to 1.09	E > O
Other	6	.51*	Yes	.28 to .75	
PA skills taught					
One skill	14	.77*	No	.58 to .96	*ns*
Two skills	11	.89*	Yes	.72 to 1.05	
Three or more skills	3	.93*	No	.52 to 1.33	
Blend and segment only	6	.85*	Yes	.54 to 1.16	*ns*
Three or more skills	3	.93*	No	.52 to 1.33	
Letter use					
Letters manipulated	17	1.00*	Yes	.85 to 1.15	L > NoL
Letters not manipulated	11	.57*	No	.37 to .76	
Delivery unit					
Individual child	8	1.00*	No	.71 to 1.28	I = S > C
Small groups	15	.94*	Yes	.78 to 1.10	
Classrooms	5	.56*	No	.33 to .78	
Length of instruction					
1 to 4.5 hours	0	—	—	—	
5 to 9.3 hours	8	1.13*	Yes	.86 to 1.39	*ns*
10 to 18 hours	8	.91*	No	.73 to 1.10	
20 to 75 hours	9	.75*	Yes	.50 to 1.01	
Instructor					
Classroom teachers	8	.74*	No	.58 to .91	*ns*
Researchers and others	20	.96*	No	.79 to 1.14	
Phonemic awareness outcomes					
Letter use					
Letters manipulated	25	1.11*	No	.99 to 1.23	L > NoL
Letter not manipulated	32	.83*	No	.73 to .92	

Note. d = mean effect size; Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous; PA = phonemic awareness; *ns* = results were not statistically significant.

a Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters/numbers of levels.

* Indicates that effect size was statistically greater than zero at $p < .05$.

instruction involving multiple skills taught with letters, the instruction was individualized, it tended to be lengthy (over 19 hours), and researchers or computers rather than classroom teachers were more often the instructors. Effects of these moderators were reexamined with RD students removed from the comparisons. Effect sizes were reanalyzed on PA and spelling outcomes because effect sizes were statistically smaller for RD students than for the other two reader groups on these outcomes. Effect sizes on reading outcomes were not reanalyzed because RD students did not show smaller effects than both of the other groups. Moderator effects involving computers were not reanalyzed because there were too few comparisons. For PA outcomes, there was only one moderator whose findings changed statistically, that involving use of letters. For spelling outcomes, several findings changed. Results are reported in Table 5 and discussed below.

Characteristics of PA instruction. Studies varied in the particular PA skills that were taught. Most commonly, students learned to identify, categorize, blend, segment, or delete phonemes, or manipulate onset-rime units. However, there were too few studies examining single types of phoneme manipulations to permit a meta-analysis.

Some studies taught single PA skills while others taught two skills or multiple skills. From statistical tests of PA outcomes in Table 2, it is apparent that focusing instruction on one or two skills was more effective for teaching PA than focusing on multiple skills (d = 1.16 for one vs. d = 1.03 for two vs. d = 0.70 for multiple). Transfer to reading was statistically greater, in fact twice as great when PA instruction focused on one (d = 0.71) or two (d = 0.79) PA skills than when multiple skills were taught (d = 0.27). The advantage of focused over multiple-skills instruction for reading persisted statistically at the follow-up test, especially for the two-skill focus, which produced statistically larger effects than the one-skill focus. These findings suggest that PA instruction focusing on one or two skills exerts a bigger impact on PA and reading outcomes than PA instruction teaching multiple PA skills.

As evident in Table 4, spelling effect sizes for focused and multiple-skills instruction showed the same pattern. However, the statistically smaller effect size in the multiple condition most likely occurred because disabled readers were overrepresented in this category and their spelling did not benefit from PA instruction (see above). When RD comparisons were removed, effect sizes became very similar and did not differ statistically (see Table 5) although there were only three comparisons in the multiple set.

Of the various combinations of phoneme manipulations that might be taught, two are thought to play a central role in learning to read and spell words. Blending phonemes helps children decode unfamiliar words. Segmenting words into phonemes helps children spell unfamiliar words and also retain spellings in memory. A number of studies examined PA instruction that taught children to blend and segment phonemes. To assess its value, the effect size for this treatment was compared to the effect size for the multiple-skills treatment. As evident in Table 2, neither approach was more effective than the other for teaching PA. However, Table 3 shows that teaching students to blend and segment statistically benefited their reading (d = 0.67) more than a multiple-skills approach did (d = 0.27). As shown in Table 4, the blending and segmenting treatment also produced a statistically larger effect on spelling performance (d = 0.79) than the multiple-skills treatment (d = 0.23). However, when comparisons involving disabled readers were removed, the difference favoring blend-and-segment instruction on spelling disappeared statistically although only three comparisons remained in the multiple-skills set (see Table 5). From these findings, we conclude that blend-and-segment instruction benefited children's reading more than multiple-skills instruction did.

Another feature of PA instruction expected to be important involved the use of letters to teach PA. In some studies, children learned to segment words into phonemes by selecting plastic letters for the sounds they detected, whereas in other studies, children only spoke the sounds or they represented the sounds with unmarked tokens. Letters were expected to improve children's acquisition of PA because they provide concrete, lasting symbols for sounds, which are short-lived and harder to grasp. From PA outcomes in Table 2, it is apparent that children taught with letters did not acquire stronger PA (d = 0.89) than children taught without letters (d = 0.82). The two effect sizes did not differ statistically. However, all but one of the RD comparisons fell in the letter condition. Because RD students gained relatively less PA than non-RD students (see Table 2), their presence may have suppressed the effect size for letters. Removal of RD students from the set revealed a statistically significant difference favoring letter instruction (d = 1.11) over no letters (d = 0.83) (see Table 5). These findings indicate that teaching non-RD children to manipulate phonemes using letters is more effective for acquiring PA than teaching them without letters.

Teaching PA with letters was also expected to promote superior transfer to reading and spelling than teaching PA without letters because reading and spelling require knowledge of phoneme-letter linkage. From Table 3, it is apparent that PA instruction with letters created statistically greater effect sizes on reading. In fact, the effect size was almost twice as large as the effect size without letters, (d = 0.67 vs. 0.38). The same statistically significant pattern persisted at the follow-up test as well

(d = 0.59 vs. 0.36). Likewise, letters benefited spelling more than no letters, with the effect size statistically greater, in fact almost twice as great (d =0.61 vs. 0.34). These findings reveal that PA instruction makes a stronger contribution to reading and spelling performance when the instruction includes teaching children to manipulate phonemes with letters than when instruction is limited to speech.

It is commonly believed that tutoring is the most effective way to deliver instruction because tutors can tailor their teaching to individual needs. However, results of our analysis did not support this. Table 2 shows that PA was taught most effectively in small groups where the effect size (d = 1.38) was very large. It was statistically greater, in fact, over twice the size of effects for tutoring (d = 0.60) and for classrooms (d = 0.67). The effectiveness of small groups may arise from enhanced attention, or social motivation to achieve, or observational learning opportunities.

Phonemic awareness instruction delivered to small groups also boosted reading and spelling performance more than tutoring or classrooms, as indicated by statistical tests. Effect sizes on reading outcomes for small groups were d = 0.81 on the immediate posttest and d = 0.83 on the follow-up posttest. In contrast, effect sizes for children tutored or taught in classrooms were statistically smaller and ranged from d = 0.30 to 0.45 on the immediate and delayed posttests. On spelling outcomes, small-group instruction produced a statistically larger effect size than tutoring did (see Table 4). However, when RD comparisons were removed, the difference was no longer statistically significant (see Table 5).

These findings are only partially consistent with those of B&I (1999). They found that small groups were more effective than tutoring for teaching PA as we did. However, on reading outcomes, they found larger effects for tutoring than for small groups in contrast to our findings.

The possibility that small-group effect sizes might be inflated for statistical reasons was considered. Some studies used group as the unit of analysis to calculate effect sizes. This may have inflated values of d because the standard deviations of group means are smaller than the standard deviations of individual scores. However, there were only five studies that used group as the statistical unit of analysis, and these contributed only six cases (13%) to the total of 45 cases in which children were taught in small groups. The small number of instances rules out statistical inflation as the explanation for small-group effect sizes.

It is common wisdom that greater time spent instructing students yields superior learning. The length of time allocated for PA instruction varied across studies from 1 hour to 75 hours. Comparisons were grouped into four time blocks to determine whether longer proved better. Analysis of PA outcomes revealed that effect sizes were statistically larger for the two middle time periods lasting from 5 to 9.3 hours (d = 1.37) and from 10 to 18 hours (d = 1.14). Periods that were either shorter or longer than this were statistically less effective for teaching PA, in fact, only half as effective (d = 0.61 and 0.65). On reading outcomes, programs that were long-lasting yielded a statistically smaller effect size than shorter programs as shown in Table 3. Effect sizes for the three shorter time blocks did not differ statistically. The same pattern was evident on spelling outcomes. However, RD students were among those who received instruction lasting the longest. When they were removed from the set of comparisons, the statistically significant difference disappeared (see Table 5), indicating that teaching time made little difference on spelling outcomes among the nondisabled readers. These findings suggest that PA instruction does not need to be lengthy to exert its strongest effect on reading and spelling.

Classroom teachers are the primary purveyors of reading instruction, so it is important to verify that they can teach PA effectively. Results in Table 2 show that classroom teachers produced a large effect size on PA outcomes (d = 0.78), although not quite as large statistically as that produced by researchers (d = 0.94). PA instruction delivered by teachers transferred to reading and spelling, as indicated by effect sizes statistically greater than zero. In the case of reading outcomes, the effect size associated with classroom teachers was statistically smaller (d = 0.41) than the effect size of researchers (d = 0.64). However, it is important to note that in most studies, the instructor did not intervene and help children apply their PA skills in the reading transfer tasks. If transfer occurred, it was unassisted. This contrasts with normal classroom operations where teachers not only teach phonemic awareness but also teach children how to apply it in their reading and provide practice doing this. Under these circumstances, much more transfer to reading would be expected.

In the case of spelling outcomes, Table 4 reveals that classroom teachers produced statistically greater effect sizes than researchers (d = 0.74 vs. 0.51). However, the researcher effect size may have been depressed by the disproportionate presence of disabled readers in this category. When disabled readers were removed from the database, the effect sizes did not differ statistically (see Table 5).

There were only seven studies that used computers to teach PA. Ten treatment-control comparisons were derived from these studies. From Table 2, it is apparent that computers produced a statistically significant and moderately strong effect size on the acquisition of PA (d =

0.66), although it was statistically less than the effect size for other forms of instruction ($d = 0.89$). PA computer instruction transferred and improved children's reading performance statistically ($d = 0.33$), but computers did not improve reading as much statistically as other forms of PA instruction did ($d = 0.55$). In contrast to positive effects on reading, PA computer instruction exerted no statistically significant effect on spelling outcomes ($d = 0.09$). One factor suppressing its impact was the presence of disabled readers who contributed 63% of the spelling effect sizes to the computer comparisons. From these findings we conclude that computers are effective for teaching PA and for promoting transfer to reading, but they may not be so effective for teaching spelling.

Design features. Design features of the studies were analyzed to query whether positive effect sizes arose primarily from strongly or weakly designed experiments. Studies varied in how students were assigned to treatment and control groups: random assignment, nonequivalent groups, matching. In the analysis of PA outcomes, effect sizes were not statistically different across these levels, ranging from $d = 0.83$ to 0.92 (see Table 2). On reading outcomes, the effect size for randomly assigned groups ($d = 0.63$) was statistically greater than the effect size for nonequivalent groups ($d = 0.40$), with matching in the middle and not statistically different from the others. The opposite was found on spelling outcomes, with nonequivalent and matched groups showing statistically larger effect sizes ($d = 0.86$ and 0.73, respectively) than random groups ($d = 0.37$). These findings confirm that larger effect sizes did not consistently arise from the weaker design involving nonequivalent groups. On the most important measure, reading, random assignment produced the strongest effect.

In some studies, fidelity checks were conducted to assess whether instructors adhered to prescribed instructional procedures. These studies revealed statistically significant effects that were moderate in size, with *d*s ranging from 0.43 to 0.66 across outcomes. However, effect sizes were statistically larger in studies not checking for fidelity, with *d*s ranging from 0.59 to 1.02 (see Tables 2, 3, and 4).

Bus and van IJzendoorn (1999) reported an unexpected finding involving the type of control group in their PA meta-analysis. When groups receiving PA instruction were compared with control groups that received no special attention, the resulting effect size on reading outcomes was statistically *smaller* than when groups receiving PA instruction were compared with control groups that received another non-PA type of instruction. One would expect the opposite to occur because Hawthorne effects should inflate the effect size in comparisons using untreated control groups. In the present study, the B&I finding was replicated. On reading outcomes, comparisons using treated controls showed statistically larger effect sizes ($d = 0.65$) than comparisons using untreated controls ($d = 0.41$). However, on the other outcomes, results did not replicate this pattern. On PA outcomes, the two types of control groups yielded about the same effect sizes, which did not differ statistically (see Table 2). On spelling outcomes, studies showed the reverse pattern. Untreated controls produced statistically larger effects than treated controls (see Table 4).

The foregoing results emerged across studies. We also compared effects within eight studies that included both treated and untreated control groups. The advantage of looking within studies is that other differences are controlled. On reading outcomes, six of the pairs exhibited larger effect sizes for untreated control groups while two showed larger effect sizes for treated controls. However, this difference fell short of statistical significance on a matched-pair *t*-test ($p > .05$). In sum, the fact that studies using untreated controls did *not* uniformly yield statistically larger effect sizes than studies using treated controls serves to challenge the commonly held belief that untreated control groups produce larger effects. It is not the case that Hawthorne effects prevail. Other factors appear to influence outcomes as well.

Studies conducted with larger samples of participants provide a stronger test of hypotheses than small-sample studies. The comparisons in our database varied in sample size from 9 students to 383 students. In our meta-analysis, the comparisons were divided into four blocks whose mean effect sizes were compared statistically. Outcomes reported in Tables 2–4 reveal that statistically larger effect sizes tended to occur in the smaller samples, whereas the smallest effect sizes occurred in the largest samples. This is consistent with meta-analytic findings in general (Johnson & Eagley, in press). The fact that effect sizes were statistically greater than zero even in the largest samples, however, shows that the positive effects of PA instruction did not arise primarily from weaker studies with small samples.

In Troia's (1999) recent critique of phonemic awareness instruction studies, he identified several criteria to assess methodological rigor and applied these criteria to 39 PA studies, of which 28 were in our database. (Troia's other studies did not include reading as an outcome so they were not in our set.) Troia devised two measures and applied them to evaluate the internal validity separately from the external validity of studies: the percentage of criteria met and the number of critical flaws. He also ranked the studies to indicate their overall methodological rigor. We adopted Troia's summary ratings and examined their relationship to effect sizes. Our purpose was to address the possibility that effects of PA instruction emerged primarily from studies that were the least rigorous. We grouped comparisons into blocks of three or four in order to reveal effect sizes at the various levels of rigor and to compare them statistically.

Our findings are reported in Table 6 for PA outcomes and Table 7 for reading outcomes. Both tables reveal that effect sizes were statistically greater than zero across all blocks on all five measures. This shows that statistically significant effect sizes were not limited to the weakest studies.

In Table 6 reporting effects of PA instruction on PA outcomes, inspection of *d* values reveals that across all five measures the largest effect sizes occurred for the blocks associated with the most rigor. Statistical tests showed that studies in the most rigorous block had significantly larger effect sizes than studies in the other blocks in three of the analyses (i.e., the two critical flaws analyses, the overall ranking analysis). These findings indicate that the best designed studies produced the largest effect sizes on the acquisition of PA.

In Table 7 reporting effect sizes for reading outcomes, the same pattern is evident but is not quite as strong. In four analyses, the effect size for the most rigorous block was large and among the largest effect sizes in the set. Statistical tests revealed that the most rigorous block produced a statistically larger effect size than the other less rigorous blocks in two analyses (i.e., internal validity flaws and overall ranking). These results indicate that the better designed studies produced stronger effects on reading than the weaker studies.

In sum, although Troia (1999) found fault with PA studies, his findings do not undermine claims about the effectiveness of PA instruction for helping children learn to read. Troia's concluding plea, that researchers maintain high standards in designing their studies, is supported by our findings that show that researchers stand a better chance of obtaining sizeable effects when they design strong studies than when they design weak studies threatened by violations to internal and external validity.

Table 6 Phonemic Awareness outcomes: Mean effect sizes (*d*) associated with Troia's indicators of methodological rigor and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$

Variables and levels		Number of cases	Mean *d*	Homogeneous	Contrasts[a]
Internal validity					
% of criteria met					
27–40%	(1)	10	.67*	Yes	2 = 4 > 1
47%	(2)	5	1.35*	No	4 > 3
53%	(3)	14	.95*	No	2 = 3
59–82%	(4)	14	1.66*	No	
Critical flaws					
1–2		18	1.63*	No	1 > 3 > 4
3		14	.57*	Yes	
4–5		11	.97*	No	
External validity					
% of criteria met					
47–53%	(1)	10	.92*	No	4 > 1 = 2
56–60%	(2)	14	.81*	No	3 = 2,4,1
63–67%	(3)	8	1.13*	No	
73–81%	(4)	11	1.40*	No	
Critical flaws					
0 flaws		13	1.69*	No	0 > all
1		8	.96*	No	1 = 2 = 3
2		13	.61*	Yes	
3		9	.97*	No	
Ranking					
High rigor (1–12)		15	1.56*	No	H = M > L
Middle (13–24)		11	1.40*	No	
Low (25–36)		17	.69*	Yes	

Note. d = mean effect size. Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous.
[a] Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters or numbers of levels.
* Indicates that effect size was statistically greater than zero at $p < .05$.

Table 7 Reading outcomes: Mean effect sizes (d) associated with Troia's indicators of methodological rigor and statistical tests to determine whether effect sizes were statistically greater than zero at $p < .05$, were homogeneous at $p < .05$, and differed from one another at $p < .05$

Variables and levels		Number of cases	Mean d	Homogeneous	Contrasts[a]
Internal validity					
% of criteria met					
27–40%	(1)	11	.49*	No	2 > 1
47%	(2)	15	.85*	No	4 > 1
53%	(3)	16	.63*	No	2 = 3 = 4
59–82%	(4)	14	.83*	No	1 = 3
Critical flaws					
1–2		22	.99*	No	1 > 3 = 4
3		18	.59*	Yes	
4–5		16	.56*	No	
External validity					
% of criteria met					
47–53%	(1)	16	.98*	No	1 > 2, 3
56–60%	(2)	14	.58*	Yes	1 = 4
63–67%	(3)	15	.61*	No	2 = 3 = 4
73–81%	(4)	11	.66*	No	
Critical flaws					
0 flaws		17	.90*	No	0 = 3 > 1
1		11	.51*	No	2 = all
2		17	.57*	Yes	
3		11	.92*	No	
Ranking					
High rigor (1–12)		19	1.00*	No	H > M = L
Middle (13–24)		14	.61*	Yes	
Low (25–36)		23	.58*	No	

Note. d = mean effect size; Homogeneous reports results of the statistical test of whether the set of effect sizes was homogeneous.
[a] Contrasts portray results of post hoc pairwise statistical tests comparing levels of moderator variables. Symbols are initial letters or numbers of levels.
* Indicates that effect size was statistically greater than zero at $p < .05$.

One final characteristic of studies that we examined was the year of publication. Studies were grouped into four time periods. From Tables 2 and 3, it is apparent that a spate of PA instruction studies was published between 1991 and 1994. Over twice as many studies were published during this period as during the other periods. The 1991 to 1994 studies also yielded statistically larger effect sizes on PA than earlier or later studies, and yielded statistically larger effect sizes on reading than two of the other three time periods. Why this occurred is not clear. The relationship between year of publication and spelling outcomes was not analyzed because too few comparisons occurred in two of the time blocks.

Classroom instruction in PA: Some illustrations. One question of special interest in our analysis was whether classroom teachers could teach PA effectively to their students. The database included 15 studies that utilized classroom teachers to deliver PA instruction. As reported above, the meta-analysis showed that teachers were effective in teaching phonemic awareness to students, and this instruction boosted children's reading and spelling performance. A few studies are sketched below to illustrate what kinds of PA manipulations were taught to students at various grade and reader levels, the teacher training components of these studies, how the studies were structured, and the strength of effects on outcomes.

Lundberg, Frost, and Petersen (1988) examined the effectiveness of an 8-month-long, carefully structured program for kindergartners. Twelve classroom teachers in Denmark taught children daily to attend to sounds in speech and to manipulate sounds through games and exercises that increased in difficulty as the year progressed. The program began with easy listening activities and rhyming exercises. Then kindergartners learned to segment sentences into words and to focus on the length of words in speech. Then words were analyzed into syllables. For example, children listened to a troll that spoke peculiarly, syllable by syllable, and they figured out what he said. Phoneme analysis was introduced in the third month by having children identify phonemes in initial positions of words. The teacher helped children find the sounds by stretching them, for example, "Mmmmmark," or repeating them, for example, "T-T-T-Tom." Also children practiced adding and deleting phonemes from words. In the fifth month, phoneme segmentation and blending were introduced, first with two-phoneme words followed by longer words. Many of the activities were designed for children's enjoyment and consisted of dancing, singing, and other noncompetitive social games.

Teachers completed an inservice course that provided them with theoretical background as well as videotaped examples of teaching procedures and gave them practice performing these procedures during the year prior to implementing the program. Teachers from the control group followed the regular preschool program, which emphasized social and aesthetic aspects of development.

This PA program was adapted and tested by Schneider, Kuspert, Roth, Vise, & Marx (1997) with German kindergartners. This study included two experiments and a total of 22 teachers who taught PA in the treatment conditions. Control groups received the regular kindergarten curriculum. Teacher training was less extensive in the German study, lasting only 2 months.

In both Danish and German studies, PA instruction produced large effect sizes on the acquisition of phonemic awareness, ranging from $d = 0.70$ to 0.82. Effect sizes on reading outcomes were small to moderate when measured the following year in first grade: $d = 0.19$ (Denmark); $d = 0.26$, and 0.45 (Germany).

The Danish program was adapted and tested with English-speaking kindergartners in a much smaller study by Brennan and Ireson (1997). The impact of instruction on word reading was large, with an effect size of $d = 1.17$, providing evidence that the Danish program can be used effectively in U.S. classrooms. An English translation of the program has been published (Adams, Foorman, Lundberg, & Beeler, 1998).

Whereas the Danish kindergarten program did not include letter manipulation, a program developed by Blachman, Ball, Black, and Tangel (1994) for kindergartners did include letters. They taught 10 teachers and their teaching assistants to deliver PA instruction to low-income, inner-city kindergartners. Children were taught in groups of four or five for 15–20 minutes per day, four times each week. The program lasted 11 weeks. The teachers received instruction in seven 2-hour inservice workshops during which they learned a theoretical framework, practiced instructional activities, and asked questions about how to implement the program.

A key activity in Blachman et al.'s (1994) program was the "say it and move it" procedure. Children learned to move a blank tile down a page as they pronounced each phoneme in a word. After children practiced segmenting two- and three-phoneme words in this way, letter-sound correspondences were taught, and they practiced segmenting the words with blank markers and letters. Additional segmentation activities were included such as moving markers into Elkonin boxes (i.e., horizontal row of blank squares drawn on paper) to represent phonemes in three-phoneme words. A variety of games was used to reinforce grapheme-phoneme correspondences. The control group followed a traditional kindergarten curriculum that included instruction in letter names and sounds. Results of the study were very positive. Children receiving PA instruction statistically outperformed controls on PA tasks, with an effect size of $d = 1.83$. Instruction transferred to reading, $d = 0.65$, and to spelling, $d = 0.94$.

Another program in our database was administered by teachers to small groups of older disabled readers. Williams (1980) tested the ABDs program, which taught students ages 7 to 12 years to segment and blend phonemes first in speech and then using letters. Children worked with a limited set of seven consonants and two vowels. Lessons progressed from segmenting words into syllables to segmenting words into phonemes, at first two phonemes followed by three phonemes. Then blending was applied to the same words. Children performed manipulations at first with wooden markers and letters later on. Their work blending letters was the equivalent of learning to decode, and their work segmenting with letters was equivalent to learning to spell the sounds in words. More letters were added later in the program as well as words with consonant clusters and two-syllable words. The program included various games, worksheets, and activities to teach these skills.

Teachers attended a half-day session to learn about the program, which was fully described in a manual. The 17 teachers were asked to use the program 20 minutes daily. Their instruction was closely monitored. Although there were 12 units, only a few teachers got through the entire program in the 26-week period.

Williams (1980) evaluated the ABDs program again the following year with another group of teachers. They had completed about half of the program when the school year ended. Treatment groups were compared to untreated control groups. The influence of PA instruction on students' ability to decode words and nonwords was measured at the end of instruction. Effect sizes were large, $d = 1.05$ for the first year, and $d = 0.97$ for the second year, showing that the ABDs program was highly effective at improving decoding skill in disabled readers.

A second program employed with disabled readers was the ADD program (Auditory Discrimination in Depth) developed originally by Lindamood and Lindamood (1975) to teach PA and its application in reading and spelling. The most recent version is the LIPS program (Lindamood & Lindamood, 1998). The unique feature of this program is that it teaches children to identify and monitor articulatory gestures associated with phonemes. Phoneme segmentation is difficult because phonemes are coarticulated in speech without any boundaries signaling where one phoneme ends and the next begins. One way to identify separate phonemes is to monitor the changes that occur in the mouth as one pronounces words. This involves directing attention to the position and shape of the lips and tongue. For example, there are three phonemes in *meat*, and these are reflected in three successive mouth movements: lips closing for /m/, lips opening into a smile shape for the vowel, then tongue tapping the roof of the mouth for /t/. Pictures of mouth positions and mirrors help children explore their own mouths and distinguish the sequence of phonemes in pronunciations of words.

Four studies in our database implemented the Lindamood program to teach PA (Kennedy & Backman, 1993; McGuinness, McGuiness, & Donahue, 1995; Wise, Ring, & Olson, 1999, 2000). Some used classroom teachers, and some used computers to deliver instruction. Children received extensive instruction to discover and categorize the various phonemes in English by analyzing their own mouth movements. They learned to label these sounds, for example, lip poppers (/p/, /b/), tip tappers (/t/, /d/), and scrapers (/k/, /g/). They learned to track these movements in spoken words in order to identify the separate phonemes and then to represent the phonemes with graphemes. Effect sizes on reading outcomes were variable, ranging from large, $d = 1.22$ for first graders (McGuinness et al., 1995), to small, $d = 0.15$ for older disabled readers (Wise et al., 1999).

One final program illustrates the type of PA instruction appropriate for preschoolers. Although taught by researchers, it could be delivered effectively by preschool teachers or parents. In a program called Sound Foundations, Byrne and Fielding-Barnsley (1991) taught a single PA skill, phoneme identity. Children learned to recognize instances of the same sound in both initial and final positions across different words. They were shown several large posters covered with pictures of objects, and they tried to pick out those having a specified beginning sound, for example, *sea, seal, sailor, sand*, or those having the same ending sound, for example, *bus, hippopotamus, horse, octopus*. Also, children were shown an array of pictures on worksheets or cards, and they selected those having targeted sounds. In each session, one phoneme in one position was taught. The letter representing that phoneme was introduced as well.

In this study, 4-year-old preschoolers received either the PA instruction described above or a control treatment that focused on story reading and semantic activities using the same posters and worksheets. Children were taught in groups of four to six for one 30-minute lesson per week lasting 12 weeks. At the end of instruction, children in the PA group were able to identify more initial and final phonemes in words than control students, not only sounds they had practiced but also unpracticed sounds, indicating that phoneme identity skill transferred to untaught phonemes. Students were also given a simplified word reading task in which they were shown a written word and asked which of two spoken choices matched the word (e.g., "Does this say *sat* or *mat*?). PA students read more words than control students, indicating that PA instruction improved preschoolers' rudimentary word recognition skill. Not only immediate but also long-term benefits of PA instruction on reading were observed.

Discussion

In our meta-analysis, we limited attention to controlled experiments because these provide the strongest scientific evidence supporting causal inferences about the impact of PA instruction on learning to read. Findings were uniformly positive. The benefits of PA instruction were replicated multiple times across experiments and thus provided solid support for the claim that PA instruction is more effective than alternative forms of instruction or no instruction in teaching PA and in helping children acquire reading and spelling skills. In addition, effects of PA instruction were found to be greater under some circumstances than others.

These findings raise various issues that need to be considered. Some of the findings were unexpected and require explanation. Several findings carry important implications for practice. Also findings point to questions needing further attention from researchers. These matters are taken up in the discussion.

Issues involving student characteristics. The studies in our database included investigations of children at risk for future reading problems. Contrary to the common view that the criteria for identifying at-risk readers include being economically disadvantaged, authors of the studies did not uniformly require them to be low in SES. In fact, of the cases investigating at-risk readers, only 27% were low in SES while 37% were middle to high SES, and the SES of the remainder was not specified. At risk was defined by low phonemic awareness in 77% of the cases. In defense of this decision, research findings show that one of the two best predictors of future reading success is phonemic awareness (Share et al., 1984), so selecting at-risk readers by measuring their PA makes sense. However, because PA instruction targeted this skill, the large effect size that was detected on PA outcomes in at-risk readers may be less surprising.

One intriguing result involving at-risk readers was that the effect of PA instruction on reading was substantially larger on the follow-up posttest ($d = 1.33$) than on the immediate posttest ($d = 0.86$) (see Table 3). This contrasted with the more typical pattern of decline in effect size with the passage of time after training ended. One reason for the increase may be that different comparisons contributed effect sizes at the two test points: 27 on the immediate posttest versus 15 on the delayed posttest. To check on this, we limited attention to the 12 comparisons that assessed reading both immediately after training and after a delay. We found that most of these comparisons (i.e., 9 out of 12) showed the same pattern of larger effects on the delayed posttest than on the immediate posttest, a proportion that exceeds chance statistically at $p < .02$ and rules out this explanation. The likely explanation for the increase is that students in the at-risk studies were preschoolers, kindergartners, or first graders with low PA and limited reading skills when training began. It took time following training for their reading skills to develop and gain the maximum benefit from PA instruction.

Results of the meta-analysis evaluating effects of SES showed that PA instruction benefited low-SES children as much as middle-to-high-SES children in acquiring PA skills. This runs counter to Dressman (1999), who argues that low-SES children exhibit low PA in research studies because their phonological systems differ from that of testers and because they suffer from inhibition when tested by sociolinguistically foreign researchers—for example, students speaking a nonstandard dialect and testers speaking Standard English. Dressman bases his claim on studies showing that low-SES children perform more poorly on PA tests than middle-class children. However, he ignores evidence examining how much low-SES children *gain* in PA when they receive instruction. According to our findings, low-SES children can benefit as much from instruction as middle-to-high-SES children, despite being phonologically or culturally different from the instructors.

One very striking finding in the meta-analysis was that spelling skill in older disabled readers did not benefit at all from PA instruction in contrast to that of at-risk and normally developing readers. Various reasons for this can be entertained. Other studies have found that disabled readers have special difficulty learning to spell (Bruck, 1993). Even though their reading might be remediated, they still display spelling problems as adults. Perhaps phonological processing difficulties associated with being RD (i.e., difficulty segmenting words into phonemes, difficulty reading nonwords) make spelling especially hard to learn (Rack, Snowling, & Olson, 1992). Alternatively, perhaps PA instruction failed to help older disabled readers with their spelling because the types of words that are spelled in higher grades require knowledge of syllabic and morphemic spelling patterns rather than phoneme-grapheme knowledge. The contribution of PA instruction to gains in spelling may be limited to words that are phonemically transparent. We interpret findings to indicate that remediating spelling in disabled readers is especially difficult and requires an approach that targets spelling explicitly.

Issues involving instructional features. According to our findings, children who received instruction that focused on only one or two PA skills—for example, segmenting words into phonemes, or segmenting and blending phonemes—exhibited stronger PA and stronger transfer to reading than children who were taught three or more PA skills. Various explanations might account for the difference. Perhaps when instruction focused on one or two skills, more students mastered the skills that were taught. Perhaps teaching multiple skills impaired the attainment of phonemic insight—that is, the idea that words are made up of phonemes. Children may have become confused about the underlying principle as they moved from one skill to the next, first breaking words into sounds, then blending sounds into words, then taking sounds out of words to say new words. Clarifying whether and why multiple-skills instruction might limit children's ability to acquire PA and apply it in reading needs further study. However, given our findings, it may be prudent for teachers who are using multiple-skills programs to teach one skill at a time until it is mastered before moving on to the next, and to teach students how each skill applies in reading or spelling tasks as soon as it is taught.

More important than the number of PA skills to teach perhaps is the issue of which skills should be taught. In all of the studies, children were given PA instruction that was considered appropriate for their level of literacy development. The manipulations taught to preschoolers were much easier than the manipulations taught to older students. Factors making PA tasks easy or difficult include the type of manipulation applied to phonemes, the number and phonological properties of phonemes in the words manipulated, whether the words are real or nonwords, and whether letters are included (Stahl & Murray, 1994). The following tasks are ordered from easy (1) to difficult (6) based on findings of Schatschneider, Francis, Foorman, Fletcher, & Mehta (1999):

1. First sound comparison: identifying the names of pictures beginning with the same sound
2. Blending onset-rime units into real words
3. Blending phonemes into real words
4. Deleting a phoneme and saying the word that remains
5. Segmenting words into phonemes
6. Blending phonemes into nonwords.

In deciding which PA manipulations to teach, teachers need to take into account not only task difficulty but also how students are expected to apply the PA skill being taught. The reason to teach first sound comparisons is to draw children's attention to the fact that words have sounds as well as meanings. A reason to teach phoneme segmentation is to help beginners generate more complete spellings of words. The reason to teach phoneme blending is to help students combine letter sounds to decode words. Teaching PA effectively includes teaching the applications as well as the skill.

Our meta-analysis showed that PA instruction improved children's reading and spelling acquisition. However, the opposite causal relationship is also supported by studies showing that children acquire phonemic awareness in the course of learning to read and spell, even though they are not taught PA explicitly. The process of learning letter-sound relations and how to use them to read and spell enhances children's ability to manipulate phonemes. Studies show that people who do not learn to read in an alphabetic system do not develop much PA (Mann, 1987; Morais et al., 1987; Read et al., 1987). In the studies we analyzed, children in the control groups who received regular literacy instruction in school made some gains on PA tests. However, effect sizes showed that controls did not gain as much PA as children who received explicit PA instruction. Thus, we cannot assume that the full extent of PA needed to facilitate reading will be acquired incidentally in the course of learning to read. It must receive special attention.

One surprising finding in our analysis involved the relationship between instructional time and outcomes. Effect sizes were larger when PA instruction lasted between 5 and 18 hours than when instructional time was shorter or longer. This is a potentially important finding for schools where time is a precious resource. However, caution is needed in interpreting this result because extenuating circumstances may explain the finding. Perhaps the goals of instruction were more complex in programs lasting longer. Perhaps the students receiving instruction were harder to teach. Perhaps spending many hours in PA instruction deprived students of the reading instruction benefiting control groups. Alternatively, perhaps shorter instruction was better. The value of PA instruction may have been to help children achieve basic alphabetic insight. Going beyond this by adding further nuances or complexities may have eroded learning by producing confusion or boredom. These are some of the possible reasons why longer teaching sessions might have produced smaller effect sizes.

At this point, it is premature to draw conclusions about how long instruction in PA needs to last to be effective. Very likely the answer depends on the goals of instruction, how many different PA skills are taught, whether letters are included, how much or how little the learners already know about PA when they begin, whether they are disabled readers, what type of provision is made for facilitating transfer to reading and spelling, and so forth. Also, individual children will differ in how much instructional time they need to acquire PA. This makes it important to tailor instructional time to student learning by assessing who has and who has not acquired the skills being taught. Children who are still having trouble should continue PA instruction while those who have learned the skills should move on to other reading and writing instruction.

Not only the total instructional time but also the length of single teaching sessions must be decided. In our database, the average length of sessions was 25 minutes. Few sessions lasted over 30 minutes, and these tended to occur with older disabled readers, not with younger children. This suggests that sessions should probably not exceed 30 minutes in length.

Although PA instruction might be taught in many ways, one might wonder what the conditions for optimal learning might be. One way to construct a tentative answer is to interpret the effect sizes associated with moderator variables in our analysis as indicating which properties might produce the greatest gains in learning to read. Taking this approach, we find that the most effective circumstances may be teaching one or two PA skills with letters, especially blending and segmenting, to small groups of English-speaking at-risk students or preschoolers for 5 to 18 hours. Although this answer is speculative, it summarizes findings and is suggestive.

PA instruction benefited reading comprehension. Some educators have asserted that effects of phonemic awareness instruction are very limited. According to Coles (2000) and Krashen (2000), instruction helps children decode nonsense words but exerts no effect on their reading comprehension. However, our analysis evaluated a greater number of studies than these authors and found that PA instruction did improve students' reading comprehension (see Table 3). There were 10 studies yielding 20 treatment-control group comparisons that assessed reading comprehension as an outcome (Bradley & Bryant, 1985; Defior & Tudela, 1994; Hatcher et al., 1994; Korkman & Peltomaa, 1993; Kozminsky & Kozminsky, 1995; Lie, 1991; Uhry & Shepherd, 1993; Weiner, 1994; Wise et al., 1999, in press). Inspection of effect sizes revealed that the majority ranged from moderate to large in size:

Size of effect	Range of *d*	Number of comparisons
Large	0.72 to 1.69	6
Moderate	0.41 to 0.62	6

Small	0.14 to 0.29	3
Zero or negative	-0.66 to 0.08	5

Effect sizes close to zero tended to occur in comparisons involving disabled readers. These findings serve to refute claims that PA instruction does not benefit reading comprehension in beginning readers.

PA instruction might be expected to exert a smaller influence on reading comprehension than on word reading because the task of reading, understanding, and remembering information in text involves multiple processes. Not only must students read the words but also they must construct meaning across the words and sentences. However, reading comprehension tests at the beginning levels consist of short sentences and conceptually simple ideas. The primary determiner of success is whether the words can be read. This suggests that effect sizes might be similar on comprehension and word level tests. Results of the meta-analysis provided some support for both possibilities. The effect size on reading comprehension was smaller ($d = 0.34$) than effect sizes on experimenter-devised tests of word reading ($d = 0.61$) and pseudoword reading ($d = 0.56$), but it was similar to the effect size on standardized tests of word reading ($d = 0.32$). Reading comprehension was measured primarily with standardized tests.

Issues involving design features. One finding of possible concern is that studies that checked on instructors' fidelity to instructional procedures exhibited statistically smaller effect sizes on all three outcomes than studies not assessing fidelity to instruction. To explain why this might have occurred, we examined other characteristics of the studies that were coded for fidelity to see whether levels of other moderator variables were distributed asymmetrically across the two levels of the fidelity moderator. Chi-square tests were applied to identify whether observed distributions deviated statistically from expected distributions ($p < .05$).

The following levels of moderator variables were overrepresented in studies that assessed fidelity compared to studies that did not assess fidelity: classroom as the instructional unit, teachers as instructors, the longest block of instructional time, the largest sample sizes. Interestingly, all of these properties produced statistically smaller effect sizes than contrasting levels of their moderators, possibly explaining why effect sizes were smaller for studies checking fidelity. It is noteworthy that fidelity checking was relatively more common in studies using larger samples, both of which are properties of better designed studies.

There were several other levels of moderators that were overrepresented in studies that did not assess fidelity: instruction in one PA skill, computers and researchers as instructors, tutoring, the shortest block of instructional time, the smallest sample sizes. Several of these characteristics produced larger effect sizes than contrasting levels of their moderators. This may help to explain why the effect size was larger for studies not checking fidelity.

Another factor underlying fidelity effect sizes may be researcher beliefs. The above analyses revealed that studies involving teachers and classrooms were more likely to be checked for fidelity than studies involving researchers. Rarely were studies involving computers checked for fidelity. Reasons for this may be that the instructional procedures employed by computers and researchers are thought to be more controlled and constrainted than those in classrooms. Researchers commonly work from scripts in delivering instruction. Computers are programed to deliver instruction uniformly. Thus, researchers may believe that there is less need to check fidelity under these circumstances.

Important but neglected moderators. Although the present meta-analysis examined effects of several moderator variables, it was not possible to examine all those thought to be important. One of the variables neglected was dialect. This was because none of the studies paid attention to this variable. Regional differences at the phonemic level of language are likely to be important. For example, vowel phoneme categories differ across the U.S. Vowels in the three words *marry*, *Mary*, and *merry* are identical in some regions but different in other regions. Another dialectic difference involves preserving or deleting the final consonants in words, for example, past tense markers such as the /t/ in *looked*. Such differences may make specific PA programs more suitable in some regions than others. More research on the impact of dialectic variations on PA learning is needed. The fact that there are regional phonemic variations means that teachers implementing PA programs need to determine whether their students' dialects conform to or deviate from the phonological distinctions that are taught in the programs they are using. Ignoring deviations is likely to undermine the effectiveness and credibility of the instruction.

Another variable related to students' phonological systems but neglected in our analysis is whether English is the first or second language of students. The problem here is that some phonemes in English are not phonemes in ESL students' first language. For example, the initial sounds in *chop* and *shop* are articulated differently. To an English speaker, they are also different phonemes, because substituting one for the other signals a different word. However, to a speaker of Spanish, the two initial sounds are the same phoneme /c/. The change in articulation does not signal a different word in Spanish. The speaker either fails to notice the difference or perceives it as a slightly different way of pronouncing the same word.

Another example is when Chinese and Japanese speakers process /l/ and /r/ as the same phoneme in English words. Teachers need to be aware of this source of difficulty when they teach PA to ESL students.

Comparison to other instruction effect sizes. Findings of our meta-analysis revealed that PA instruction helped children learn to read and spell. The overall effect sizes were moderate, with $d = 0.53$ for reading and $d = 0.59$ for spelling. One might wonder how these effect sizes compare to effects of other types of instruction aimed at improving reading ability. The National Reading Panel (2000) conducted some additional meta-analyses. One involved experiments comparing systematic phonics instruction to nonsystematic or nonphonics instruction. The phonics database included 38 studies yielding 66 treatment-control group comparisons. None of these studies appeared in the present PA database. The overall effect size was $d = 0.41$ on a composite measure that included reading and spelling. This value is within the moderate range but slightly smaller than the effect sizes found for PA instruction. In the NRP meta-analysis of instruction involving guided oral reading and repeated reading taught in 14 studies, the overall mean effect size was $d = 0.41$ on a composite outcome that included comprehension, fluency, and word recognition. This effect size is moderate but also slightly smaller than that for PA.

In a meta-analysis of studies evaluating the effects of instruction in question asking on reading comprehension (Rosenshine, Meister, & Chapman, 1996), the effect size on standardized tests was $d = 0.36$ (13 studies) and on experimenter-devised tests was $d = 0.87$ (16 studies). In the present PA analysis, the effect size on reading comprehension tests, most of which were standardized, was similar, $d = 0.34$ (20 comparisons). From these findings we conclude that effect sizes produced by PA instruction were comparable to effect sizes produced by other types of interventions on reading outcomes. Thus, we can infer that the gains expected by including PA instruction in the reading curriculum are as great as the gains offered by other types of reading instruction regarded as valuable.

Classroom teachers teaching PA. One question of special interest in our analysis was whether classroom teachers can teach PA effectively to their students. Findings were resoundingly positive, as indicated by moderate to large effect sizes. Examination of studies illustrating how classroom instruction was conducted revealed that the amount of inservice instruction given to teachers was not unreasonable. Several of the programs investigated are available commercially, including *Sound Foundations* (Byrne & Fielding-Barnsley, 1991), *Ladders to Literacy: A Preschool Activity Book* (Notari-Syverson, O'Connor, & Vadasy, 1998), *Ladders to Literacy: A Kindergarten Book* (O'Connor, Notari-Syverson, & Vadasy, 1998b), *Road to the Code* (Blachman, Ball, Black, & Tangel, 2000), *Phonemic Awareness in Young Children* (Adams et al., 1998), and *The LIPS Program* (Lindamood & Lindamood, 1998).

In planning PA instruction for students, teachers need to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. At the start of kindergarten, most children will be nonreaders and will have little phonemic awareness, so PA instruction should benefit everyone. In first grade, some children will be reading and spelling already while others may know only a few letters and have no reading skill. The non-readers will need much more PA and letter instruction than those already reading. Among the readers in first and second grades, there will be variation in how well children can segment and blend with letters. The best approach is for first- and second-grade teachers to assess students' PA before they begin teaching PA. This will indicate which children need the instruction and which do not, or which children need to be taught rudimentary levels of PA, for example, segmenting initial sounds in words, and which are ready for more advanced levels involving segmenting or blending with letters. Of course, it is better to err on one side than the other: teaching PA to students who already know it is less harmful than failing to teach PA to students who lack any PA.

In the rush to teach phonemic awareness, teachers should not overlook the need to teach letters as well. Our analysis showed that PA instruction was more effective when it was taught with letters. Using letters to manipulate phonemes helps children make the transfer to reading and writing. However, teaching children all the letters of the alphabet is not easy, particularly when they come to school knowing few of them. There are 52 capital and lowercase letter shapes, and names, as well as associated sounds to learn. The shapes of many letters are similar and hence confusing. Letter learning requires retaining shapes, names, and sounds in memory and overlearning them so that letters can be processed automatically in reading and writing words (Adams, 1990). Thus, to ensure that PA instruction is effective, it needs to include instruction in graphemes and instruction in the connections between graphemes and phonemes to read and spell words.

In addition to teaching PA skills with letters, it is important for teachers to help children apply the PA skills taught in reading and writing tasks. In most of the studies we reviewed, application was not taught. However, Cunningham (1990) did examine application effects. First graders in one group were taught to segment and blend and in addition were shown how to apply these skills in reading words. Another group received the same PA instruction but not instruction in application. Effect sizes on reading outcomes were much larger when first graders received the application instruction than when they did not. Because application was not taught in most studies, results of our meta-analysis may underestimate the magnitude of effects that would result if children received explicit instruction and practice in applying PA skills in their reading and writing.

Various approaches to beginning reading instruction may provide at least some phonemic awareness instruction although PA may not be taught systematically or thoroughly enough to maximize its contribution to reading and writing. Whole language instruction in which beginners invent spellings when they write offers a form of PA instruction. One study in our database (Castle et al., 1994) examined the effects of providing PA instruction to children who were receiving whole language reading and spelling in their classrooms. PA instruction improved these children's reading over that of two control groups, one receiving whole language instruction plus word meaning instruction ($d = 1.06$) and one receiving whole language instruction but no extra instruction ($d = 1.09$). Effect sizes were even larger on spelling outcomes. This indicates that adding systematic instruction in phonemic segmentation and blending to whole language programs holds promise of improving their effectiveness.

In another study, Iversen and Tunmer (1993) added PA instruction to the standard form of Reading Recovery (RR), which teaches phonemic awareness mainly through spelling instruction (Clay, 1985). Results showed that the RR plus PA treatment group performed similarly to the RR control group on reading and spelling outcomes when they exited the program, but the group receiving PA instruction took statistically less time to reach the exit criterion. Because saving time means that RR teachers can help more students profit from this tutoring program during a school year, these findings indicate that adding PA instruction stands to improve the program's effectiveness.

As more and more educators learn about the benefits of phonemic awareness instruction for helping children learn to read, there is a danger that PA will be regarded as a magic bullet, will be taught blindly in isolation ad nauseam without any connection to reading and writing, and will replace other important kinds of reading instruction and activities. Clearly, such instructional implications are not supported by our meta-analysis. Our findings indicate that teaching phonemic awareness is a means rather than an end. PA is taught not for its own sake but rather for its value in helping children understand and use the alphabetic system to read and write. Findings also indicate that a moderate amount of time rather than a huge amount may be sufficient to teach PA.

Our findings indicate that PA contributes significantly to reading and spelling acquisition, but there is much more that children need to be taught in order to become competent readers and writers. From effect sizes we can calculate the proportion of variance in reading outcomes explained by PA instruction. Overall the variance explained was 6.5%. It rose to 10% when PA was taught with letters. It rose to 28% for preschoolers. It rose to 31% for the long-term reading performance of at-risk students. Thus, even under the best of circumstances, there remained much variance in reading left to be explained by other types of instruction; for example, print awareness, letter naming and writing, decoding, sight word learning, spelling, vocabulary, and comprehension of text by listening as well as reading (Snow et al., 1998).

Effect sizes produced by PA instruction were especially large at the preschool level. This suggests the benefit of incorporating age-appropriate activities to teach PA into the preschool curriculum. However, results do not indicate that only PA should be taught. Other emergent literacy activities such as storybook listening have been found to make an important contribution as well (Bus, van IJzendoorn, & Pellegrini, 1995; Scarborough & Dobrich, 1994).

Directions for future research. Many experiments have been conducted showing that phonemic awareness instruction helps various types of children learn to read across a variety of conditions. However, there are still some questions needing further attention from researchers.

Studies by Moats (1994) and Scarborough, Ehri, Olson, and Fowler (1998) question whether teachers possess sufficient phonemic awareness themselves to teach this skill adequately on their own. However, findings of our analysis indicate that when teachers are given instruction, they can learn to teach PA adequately. Further research is needed to clarify what sort of knowledge and instruction maximizes teachers' effectiveness in teaching PA and in integrating this with beginning reading instruction. Very likely teachers need to be taught about phonemes, graphemes, and the alphabetic system, and about literacy learning processes and their course of development in beginning readers and in older disabled readers. Also they need to know about phonemic awareness, how it develops in children, which tasks are easier and harder, what techniques help children focus on phoneme-size units in words, what kinds of mistakes children commonly make, how they should be corrected, and so forth. Teaching children to invent spellings of words is one way to teach PA. Teachers may need to understand the processes children use to invent spellings, how their spellings become more complete and conventional, and how to promote this growth. An experiment by McCutchen et al. (in press) provides evidence that it is possible to deepen teachers' knowledge about phonemic awareness and thereby to improve their ability to teach reading and spelling.

One factor not addressed in the PA studies we reviewed is how motivating these programs are for students and also for teachers. It seems self-evident that when instructional techniques are not only effective but also are engaging, interesting, and motivating, they serve to promote optimal learning in children and to increase the likelihood that teachers will continue to use them once they learn the techniques. This is an issue for future research.

In the meta-analysis of instructional programs, the delivery unit was found to influence outcomes. Small-group instruction was associated with much larger effect sizes than individual or classroom instruction. However, findings were correlational, so inferring that small groups produced the greater gains is highly tentative and subject to alternative interpretations. What is needed are experiments that compare small groups to other delivery units. This would clarify whether small-group instruction is a better way to teach PA than individual or classroom instruction.

Use of computers is fast becoming a national pastime in the U.S., at home as well as at school. Younger children are acquiring facility with computers. Parents as well as teachers are in the market for effective computer programs to teach important skills to children. A few studies in our database examined whether computers delivered PA instruction effectively. Findings showed that effect sizes were statistically significant for teaching PA and its transfer to reading. However, effects were smaller than those produced by teachers or researchers. Computers were of doubtful value for promoting transfer to spelling, although this may apply only to older disabled readers. More research is needed to determine whether and how PA might be taught effectively using computers.

Many parents of preschoolers are anxious to help their children acquire the knowledge and skills they need to become successful when they enter school and begin reading instruction. However, none of the studies we reviewed used parents as instructors. Research is needed to address this gap in our knowledge. In addition to informal activities that parents might use to draw preschoolers' attention to sounds in words, the effectiveness of activities that help parents teach letters and their connection to sounds should be studied.

Results of our meta-analysis revealed the value of experimental studies for providing reliable findings that can guide instructional practice. We examined whether well-designed studies yielded stronger effect sizes than weaker designs and found that effect sizes were largest for studies that were methodologically rigorous. These findings underscore the importance of researchers taking steps to maximize the rigor of their designs by addressing as many threats to internal and external validity as is possible. Not only does this enhance confidence in the findings but also, as our meta-analysis shows, this gives researchers a better chance of detecting treatment effects when they exist.

Limitations. Several limitations of our findings and conclusions have been noted. Conclusions that PA instruction impacted outcomes more under some circumstances than under others must be regarded as tentative. These findings are not based on treatment-control comparisons within studies but are based on correlations between studies. Strong causal inferences are precluded because a third factor may lie behind the scenes and explain the relationship. For example, in our analyses of the effects of PA instruction on spelling, we found that the effect size was statistically greater for classroom teachers ($d = 0.74$) than for researchers ($d = 0.51$). However, when a third factor was considered and its effects removed, the finding changed. Specifically, most disabled readers were taught PA by researchers, not by classroom teachers. PA instruction did not improve disabled readers' ability to spell. Removing them from the analysis boosted the effect size associated with researchers to $d = 0.96$ but did not alter the effect size for teachers, and the difference was no longer statistically significant. (For other examples, compare Tables 4 and 5.) Thus, findings involving moderators are tentative and suggestive, not conclusive.

Another limitation of our analysis is that not all of our studies contributed to all of our effect sizes. This was because some studies did not measure the outcomes of interest immediately after training, which was the test point used to calculate moderator effect sizes. Also some studies failed to report moderator information needed to code studies. Studies and their codes are listed in the Appendix, so information about which studies contributed to which effect sizes is provided. Because the particular studies contributing effect sizes to moderators varied, one must be cautious in comparing the effect sizes of different moderators. For example, comparing effect sizes associated with SES to any other moderator overlooks the fact that few RD comparisons contributed effect sizes to SES in contrast to other moderators.

Limitations regarding the implications for instruction have been noted. Teaching PA is not the sole key to reading success nor does it constitute a complete beginning reading program. Many other capabilities must be taught as well. PA is not a skill to be taught in isolation. PA must be taught in conjunction with grapheme-phoneme knowledge and application of this knowledge to reading and writing. We cannot assume that teachers automatically know how to teach PA by virtue of being readers and writers themselves. Most will require instruction in how to teach PA in order to deliver it effectively to students. Not all the answers have been provided by research on PA instruction. There are many additional questions and issues awaiting investigation.

REFERENCES

Note: An asterisk marks those publications that provided data for the meta-analysis.

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

ADAMS, M., FOORMAN, B., LUNDBERG, I., & BEELER, T. (1998). *Phonemic awareness in young children: A classroom curriculum.* Baltimore, MD: Brookes.

APTHORP, H. (1998, April). *Phonological awareness training for beginning readers: A meta-analysis of reading posttests.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

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

*BARKER, T., & TORGESEN, J. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research, 13,* 89–103.

*BENTIN, S., & LESHEM, H. (1993). On the interaction between phonological awareness and reading acquisition: It's a two-way street. *Annals of Dyslexia, 43,* 125–148.

BLACHMAN, B. (2000). Phonological awareness. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 483–502). Mahwah, NJ: Erlbaum.

*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., BALL, E., BLACK, R., & TANGEL, D. (2000). *Road to the code.* Baltimore, MD: Brookes.

BLOOM, B. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 4–16.

*BRADLEY, L., & BRYANT, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301,* 419–421.

*BRADLEY, L., & BRYANT, P. (1985). *Rhyme and reason in reading and spelling* (International Academy for Research in Learning Disabilities, Monograph Series, *1,* 75–95). Ann Arbor, MI: University of Michigan Press. (This is a more complete report of Bradley & Bryant, 1983)

*BRADY, S., FOWLER, A., STONE, B., & WINBURY, N. (1994). Training phonological awareness: A study with inner-city kindergarten children. *Annals of Dyslexia, 44,* 26–59.

*BRENNAN, F., & IRESON, J. (1997). Training phonological awareness: A study to evaluate the effects of a program of metalinguistic games in kindergarten. *Reading and Writing: An Interdisciplinary Journal, 9,* 241–263.

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

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

*BUS, A. (1986). Preparatory reading instruction in kindergarten: Some comparative research into methods of auditory and auditory-visual training of phonemic analysis and blending. *Perceptual and Motor Skills, 62,* 11–24.

BUS, A., & VAN IJZENDOORN, M. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403–414.

BUS, A., VAN IJZENDOORN, M., & PELLEGRINI, A. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65,* 1–21.

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

*BYRNE, B., & FIELDING-BARNSLEY, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85,* 104–111. (This is a first follow-up to Byrne & Fielding-Barnsley, 1991)

*BYRNE, B., & FIELDING-BARNSLEY, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology , 87,* 488–503. (This is a second follow-up to Byrne & Fielding-Barnsley, 1991)

*CASTLE, J. M., RIACH, J., & NICHOLSON, T. (1994). Getting off to a better start in reading and spelling: The effects of phonemic awareness instruction within a whole language program. *Journal of Educational Psychology, 86,* 350–359.

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

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

COHEN, P., KULIK, J., & KULIK, C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19,* 237–248.

COLES, G. (2000). *Misreading reading.* Portsmouth, NH: Heinemann.

COOPER, H. (1998). *Synthesizing research.* Thousand Oaks, CA: Sage.

*CUNNINGHAM, A. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology, 50,* 429–444.

*DAVIDSON, M., & JENKINS, J. (1994). Effects of phonemic processes on word reading and spelling. *Journal of Educational Research, 87,* 148–157.

*DEFIOR, S., & TUDELA, P. (1994). Effect of phonological training on reading and writing acquisition. *Reading and Writing: An Interdisciplinary Journal, 6,* 299–320.

DRESSMAN, M. (1999). On the use and misuse of research evidence: Decoding two states' reading initiatives. *Reading Research Quarterly, 34,* 258–285.

EHRI, L. (1979). Linguistic insight: Threshold of reading acquisition. In T.G. Waller & G.E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 1, pp. 63–114). New York: Academic Press.

EHRI, L. (1980). The development of orthographic images. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 311–338). London: Academic Press.

EHRI, L. (1991). Development of the ability to read words. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research,* (Vol. 2, pp. 383–417). New York: Longman.

EHRI, L. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107–143). Hillsdale, NJ: Erlbaum.

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., & WILCE, L. (1987a). Cipher versus cue reading: An experiment in decoding acquisition. *Journal of Educational Psychology, 79,* 3–13.

*EHRI, L., & WILCE, L. (1987b). Does learning to spell help beginners learn to read words? *Reading Research Quarterly, 22,* 48–65.

*FARMER, A., NIXON, M., & WHITE, R. (1976). Sound blending and learning to read: An experimental investigation. *British Journal of Educational Psychology , 46,* 155–163.

FAWCETT, A., & NICOLSON, R. (1995). Persistence of phonological awareness deficits in older children with dyslexia. *Reading and Writing: An Interdisciplinary Journal, 7,* 361–376.

*FOX, B., & ROUTH, D. (1976). Phonemic analysis and synthesis as word-attack skills. *Journal of Educational Psychology , 68,* 70–74.

*FOX, B., & ROUTH, D. (1984). Phonemic analysis and synthesis as word attack skills: Revisited. *Journal of Educational Psychology, 76,* 1059–1064.

GLASS, G., CAHEN, L., SMITH, M., & FILBY, N. (1982). *School class size.* Beverly Hills, CA: Sage.

GRIFFITH, P. (1991). Phonemic awareness helps first graders invent spellings and third graders remember correct spellings. *Journal of Reading Behavior, 23,* 215–233.

*GROSS, J., & GARNETT, J. (1994). Preventing reading difficulties: Rhyme and alliteration in the real world. *Educational Psychology in Practice , 9,* 235–240.

*HADDOCK, M. (1976). Effects of an auditory and an auditory-visual method of blending instruction on the ability of prereaders to decode synthetic words. *Journal of Educational Psychology, 68,* 825–831.

*HATCHER, P., HULME, C., & ELLIS, A. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65,* 41–57.

*HOHN, W., & EHRI, L. (1983). Do alphabet letters help prereaders acquire phonemic segmentation skill? *Journal of Educational Psychology, 75,* 752–762.

*HURFORD, D., JOHNSTON, M., NEPOTE, P., HAMPTON, S., MOORE, S., NEAL, J., MUELLER, A., MCGEORGE, K., HUFF, L., AWAD, A., TATRO, C., JULIANO, C., & HUFFMAN, D. (1994). Early identification and remediation of phonological-processing deficits in first-grade children at risk for reading disabilities. *Journal of Learning Disabilities, 27,* 647–659.

*IVERSEN, S., & TUNMER, W. (1993). Phonological processing skills and the Reading Recovery Program. *Journal of Educational Psychology*, *85*, 112–126.

JOHNSON, B. (1989). *DSTAT: Software for the meta-analytic 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 psychology* (pp. 496–528). London: Cambridge University Press.

JUEL, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology*, *80*, 437–447.

JUEL, C., GRIFFITH, P., & GOUGH, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, *78*, 243–255.

*KENNEDY, K., & BACKMAN, J. (1993). Effectiveness of the Lindamood Auditory Discrimination in Depth program with students with learning disabilities. *Learning Disabilities Research and Practice*, *8*, 253–259.

*KORKMAN, M., & PELTOMAA, A. (1993). Preventive treatment of dyslexia by a preschool training program for children with language impairments. *Journal of Clinical Child Psychology*, *22*, 277–287.

*KOZMINSKY, L., & KOZMINSKY, E. (1995). The effects of early phonological awareness training on reading success. *Learning and Instruction*, *5*, 187–201.

KRASHEN, S. (April, 2000). Comments on the National Reading Panel. Electronic mail message posted on the National Reading Conference listserv.

LIBERMAN, I., SHANKWEILER, D., FISCHER, F., & CARTER, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201–212.

*LIE, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. *Reading Research Quarterly*, *26*, 234–250.

LINDAMOOD, C., & LINDAMOOD, P. (1975). *The ADD program: Auditory Discrimination in Depth*. Austin, TX: DLM.

LINDAMOOD, P. & LINDAMOOD, P. (1998). *The Lindamood phoneme sequencing program for reading, spelling and speech: The LIPS program*. Austin, TX: Pro-Ed.

*LOVETT, M., BARRON, R., FORBES, J., CUKSTS, B., & STEINBACH, K. (1994). Computer speech-based training of literacy skills in neurologically impaired children: A controlled evaluation. *Brain and Language*, *47*, 117–154.

*LUNDBERG, I., FROST, J., & PETERSEN, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, *23*, 263–284.

MANN, V. (1987). Phonological awareness: The role of reading experience. In P. Bertelson (Ed.), *The onset of literacy: Cognitive processes in reading acquisition* (pp. 65–92). Cambridge, MA: MIT Press.

MCCUTCHEN, D., ABBOTT, R., GREEN, L., BERETVAS, S., COX, S., POTTER, N., QUIROGA, T., & GRAY, A. (in press). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities*.

*MCGUINNESS, D., MCGUINNESS, C., & DONOHUE, J. (1995). Phonological training and the alphabet principle: Evidence for reciprocal causality. *Reading Research Quarterly*, *30*, 830–852.

MOATS, L. (1994). Knowledge of language: The missing foundation for teacher education. *Annals of Dyslexia*, *44*, 81–102.

MORAIS, J., BERTELSON, P., CARY, L., & ALEGRIA, J. (1987). Literacy training and speech segmentation. In P. Bertelson (Ed.), *The onset of literacy: Cognitive processes in reading acquisition* (pp. 45–64). Cambridge, MA: MIT Press.

*MURRAY, B. (1998). Gaining alphabetic insight: Is phoneme manipulation skill or identity knowledge causal? Journal of Educational Psychology, 90, 461–475.

NATIONAL READING PANEL. (2000). *Report of the National Reading Panel: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development Clearinghouse.

NOTARI-SYVERSON, A., O'CONNOR, R., & VADASY, P. (1998). *Ladders to literacy: A preschool activity book*. Baltimore, MD: Brookes.

*O'CONNOR, R., & JENKINS, J. (1995). Improving the generalization of sound/symbol knowledge: Teaching spelling to kindergarten children with disabilities. *The Journal of Special Education*, *29*, 255–275.

*O'CONNOR, R., JENKINS, J., & SLOCUM, T. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology*, *87*, 202–217.

*O'CONNOR, R., NOTARI-SYVERSON, A., & VADASY, P. (1996). Ladders to literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. *Exceptional Children*, *63*, 117–130.

*O'CONNOR, R., NOTARI-SYVERSON, A., & VADASY, P. (1998a). First-grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. *Learning Disabilities Research and Practice*, *13*, 43–52. (This is a follow-up to O'Connor et al., 1996)

O'CONNOR, R., NOTARI-SYVERSON, A., & VADASY, P. (1998b). *Ladders to literacy: A kindergarten activity book*. Baltimore, MD: Brookes.

*OLOFSSON, A., & LUNDBERG, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology*, 24, 35–44.

*OLOFSSON, A., & LUNDBERG, I. (1985). Evaluation of long term effects of phonemic awareness training in kindergarten: Illustrations of some methodological problems in evaluation research. *Scandinavian Journal of Psychology*, *26*, 21–34. (This is a follow-up to Olofsson & Lundberg, 1983)

PINNELL, G., LYONS, C., DEFORD, D., BRYK, A., & SELTZER M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, *29*, 9–39.

RACK, J., HULME, C., SNOWLING, M., & WIGHTMAN, J. (1994). The role of phonology in young children learning to read words: The direct-mapping hypothesis. *Journal of Experimental Child Psychology*, *57*, 42–71.

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

READ, C., ZHANG, Y., NIE, H., & DING, B. (1987). The ability to manipulate speech sounds depends on knowing alphabetic writing. In P. Bertelson (Ed.), *The onset of literacy: Cognitive processes in reading acquisition* (pp. 31–44). Cambridge, MA: MIT Press.

REITSMA, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology*, *75*, 321–339.

*REITSMA, P., & WESSELING, R. (1998). Effects of computer-assisted training of blending skills in kindergartners. *Scientific Studies of Reading*, *2*, 301–320.

ROSENSHINE, B., MEISTER, C., & CHAPMAN, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, *66*, 181–221.

*SANCHEZ, E., & RUEDA, M. (1991). Segmental awareness and dyslexia: Is it possible to learn to segment well and yet continue to read and write poorly? *Reading and Writing: An Interdisciplinary Journal*, *3*, 11–18.

SCARBOROUGH, H., & DOBRICH, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, *14*, 245–302.

SCARBOROUGH, H., EHRI, L., OLSON, R., & FOWLER, A. (1998). The fate of phonemic awareness beyond the elementary school years. *Scientific Studies of Reading*, *2*, 115–142.

SCHATSCHNEIDER, C., FRANCIS, D., FOORMAN, B., FLETCHER, J., & MEHTA, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology*, *91*, 439–449.

*SCHNEIDER, W., KUSPERT, P., ROTH, E., VISE, M., & MARX, H. (1997). Short-and long-term effects of training phonological awareness in kindergarten: Evidence from two German studies. *Journal of Experimental Child Psychology*, *66*, 311–340.

SHARE, D., JORM, A., MACLEAN, R., & MATTHEWS, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology*, *76*, 1309–1324.

SNOW, C., BURNS, M., & GRIFFIN, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

*SOLITY, J. (1996). Phonological awareness: Learning disabilities revisited? *Educational & Child Psychology*, *13*, 103–113.

STAHL, S., & MURRAY, B. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, *86*, 221–234.

*TANGEL, D., & BLACHMAN, B. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior*, *24*, 233–261.

*TORGESEN, J., MORGAN, S., & DAVIS, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology*, *84*, 364–370.

*TREIMAN, R., & BARON, J. (1983). Phonemic-analysis training helps children benefit from spelling sound rules. *Memory and Cognition*, *11*, 382–389.

TROIA, G. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly*, *34*, 28–52.

*UHRY, J., & SHEPHERD, M. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. *Reading Research Quarterly, 28,* 218–233.

*VADASY, P., JENKINS, J., ANTIL, L., WAYNE, S., & O'CONNOR, R. (1997a). Community-based early reading intervention for at-risk first graders. *Learning Disabilities Research and Practice, 12,* 29–39.

*VADASY, P., JENKINS, J., ANTIL, L., WAYNE, S., & O'CONNOR, R. (1997b). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly, 20,* 126–139.

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

VENEZKY, R. (1970). *The structure of English orthography.* The Hague: Mouton.

VENEZKY, R. (1999). *The American way of spelling.* New York: Guilford Press.

WAGNER, R. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly, 32,* 261–279.

WAGNER, R., & TORGESEN, J. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

*WARRICK, N., RUBIN, H., & ROWE-WALSH, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. *Annals of Dyslexia, 43,* 153–173.

WASIK, B., & SLAVIN, R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28,* 179–200.

*WEINER, S. (1994). Effects of phonemic awareness training on low- and middle-achieving first graders' phonemic awareness and reading ability. *Journal of Reading Behavior, 26,* 277–300.

*WILLIAMS, J. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72,* 1–15.

*WILSON, J., & FREDERICKSON, N. (1995). Phonological awareness training: An evaluation. *Educational & Child Psychology, 12,* 68–79.

*WISE, B., RING, J., & OLSON, R. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology, 72,* 271–304.

*WISE, B., RING, J., & OLSON, R. (2000). Individual differences in benefits from computer-assisted remedial reading. *Journal of Experimental Child Psychology, 77,* 197–235.

Received February 24, 2000
Revision received August 16, 2000
Accepted September 12, 2000

AUTHOR NOTE

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.

APPENDIX Studies in the PA training database, their characteristics and effect sizes

	Characteristics of training					Characteristics of participants				Features of design				Effect sizes		
Author and year, treatment vs. control	Number skills	Letters	Training unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assignment	Fidelity	*N* (Case)	*N* (Study)	PA	Reading	Spelling
Ball & Blachman, 1991													89			
01 Segment & category + let vs. Language, LS	2	yes	SmG	Other	9.33	Nor	K	Engl		R	Yes	59		1.49	0.71	0.87
02 Segment & category + let vs. No treatment	2	yes	SmG	Other	9.33	Nor	K	Engl		R	Yes	59		1.64	0.98	0.83
Barker & Torgesen, 1995													36			
03 Mult. PA on computers vs. math on computers	3+	no	Ind	Comp	13.33	AR	1st	Engl		R	No	36		0.48	0.22	
Bentin & Leshem, 1993													91			
04 Segment & category vs. Language	2	no	SmG	Other	10	AR	K	Hebr	M-H	R	No	50			4.21	
05 Segment & category vs. No treatment	2	no	SmG	Other	10	AR	K	Hebr	M-H	R	No	41			4.33	
06 Segment & category + let vs. Language	2	yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	50			2.1	
07 Segment & category + let vs. No treatment	2	yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	41			2.17	
Blachman, Ball, Black, & Tangel, 1994													159			
08 Segment & category + let vs. No treatment	2	yes	SmG	Teach	12.3	Nor	K	Engl	Lo	NE	No	159		1.83	0.65	0.94
Bradley & Bryant, 1983, 1985													65			
09 Phonetic category vs. Semantic category	1	no	Ind	Other	11.67	AR	1st	Engl		M/R	No	39			0.5	0.39
10 Phonetic category vs. No treatment	1	no	Ind	Other	11.67	AR	1st	Engl		M/R	No	26			0.86	1
11 Phonetic category + let vs. Semantic category	1	yes	Ind	Other	11.67	AR	1st	Engl		M/R	No	39			1.17	1.59
12 Phonetic category + let vs. No treatment	1	yes	Ind	Other	11.67	AR	1st	Engl		M/R	No	26			1.53	2.18
Brady, Fowler, Stone, & Winbury, 1994													42			
13 Mult. PA vs. No treatment	3+	no	Clas	Teach	18	AR	K	Engl	Lo	NE	Yes	42		0.46	0.47	0.23
Brennan & Ireson, 1997													24			
14 Segment & blend vs. No treatment	2	no	Clas	Teach	48	Nor	K	Engl	M-H	NE	Yes	24		3.92	1.17	2.17
Bus, 1986													201			
15 Segment & blend, LS vs. Preread prep., LS	2	no	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	130		0.55	0.54	
16 Segment & blend + let vs. Preread prep., LS	2	yes	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	134		0.25	0.35	
Byrne & Fielding-Barnsley, 1991, 1993, 1995													126			
17 Phonetic category + let vs. Semantic category	1	yes	SmG	Other	6	Nor	Pre	Engl	M-H	R	No	126		3.14	1.61	.34*
Castle, Riach, & Nicholson, 1994, Experiment 2													51			
18 Mult. PA + let vs. Language	3+	yes	SmG	Other	5	Nor	K	Engl	M-H	M/R	No	34		3.81	1.06	1.27
19 Mult. PA + let vs. No treatment	3+	yes	SmG	Other	5	Nor	K	Engl	M-H	M/R	No	34		2.62	1.09	1.73
Cunningham, 1990													84			
20 Segment & blend vs. Stories	2	no	SmG	Other	6	Nor	K	Engl	M-H	M/R	No	28		1.62	0.42	
21 Segment, blend, metacognitive vs. Stories	2	no	SmG	Other	6	Nor	K	Engl	M-H	M/R	No	28		2.3	0.56	
22 Segment & blend vs. Stories	2	no	SmG	Other	6	Nor	1st	Engl	M-H	M/R	No	28		0.99	0.08	
23 Segment, blend, metacognitive vs. Stories	2	no	SmG	Other	6	Nor	1st	Engl	M-H	M/R	No	28		1.27	0.51	

(continued)

APPENDIX Studies in the PA training database, their characteristics and effect sizes (continued)

	Characteristics of training					Characteristics of participants				Features of design				Effect sizes		
Author and year, treatment vs. control	Number skills	Letters	Training unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assignment	Fidelity	N (Case)	N (Study)	PA	Reading	Spelling
Davidson & Jenkins, 1994													40			
24 Segment, LS vs. Stories, LS	1	no	SmG	Other	8.33	Nor	K	Engl		NE	No	20		8	1.58	1.6
25 Blend, LS vs. Stories, LS	1	no	SmG	Other	8.33	Nor	K	Engl		NE	No	20		3.11	0.71	0.49
26 Segment & blend, LS vs. Stories, LS	2	no	SmG	Other	8.33	Nor	K	Engl		NE	No	20		3.93	1.56	1.13
Defior & Tudela, 1994													43			
27 Category + let vs. Semantic category	1	yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22			0.82	1.44
28 Category + let vs. Hand manipulation	1	yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22			0.73	1.03
29 Category vs. Semantic category	1	no	SmG	Other	30	AR	1st	Span	M-H	R	No	21			0.18	0.36
30 Category vs. Hand manipulation	1	no	SmG	Other	30	AR	1st	Span	M-H	R	No	21			0.14	0.02
Ehri & Wilce, 1987b													20			
31 Segment + let vs. LS	1	yes	Ind	Other	5.6	Nor	K	Engl	M-H	M/R	No	20		1.99	0.97	2.59
Farmer, Nixon, & White, 1976													60			
32 Blend + let vs. Label pictures	1	yes	Ind	Other		Nor	1st	Engl		R	No	20		0.78	0.96	
33 Blend + let vs. Label pictures	1	yes	Ind	Other		Nor	K	Engl		R	No	40		0.63	0.35	
Fox & Routh, 1976													40			
34 Read training with blend vs. Without blend	1	no	Ind	Other	1	Nor	Pre	Engl	M-H	R	No	20			1.61	
35 Read training with blend vs. Without blend	1	no	Ind	Other	1	AR	Pre	Engl	M-H	R	No	20			-0.1	
Fox & Routh, 1984													31			
36 Onset-rime, LS vs. No treatment, LS	1	no	SmG	Other	5	AR	K	Engl		R	No	21		0.75	-0.19	
37 Onset-rime, LS vs. No treatment, LS	1	no	SmG	Other	5	AR	K	Engl		R	No	21		1.6	3.6	
Gross & Garnett, 1994													12			
38 Category vs. No treatment	1	no	SmG	Other		AR	K	Engl	Lo	M/R	No	12			2.29*	.60*
Haddock, 1976													80			
39 Blend, LS vs. LS	1	no	Clas	Teach	2.5	Nor	Pre	Engl		NE	No	53			0.92	
40 Blend + let vs. LS	1	yes	Clas	Teach	2.5	Nor	Pre	Engl		NE	No	48			1.67	
Hatcher, Hulme, & Ellis, 1994													124			
41 Mult. PA vs. No treatment	3+	no	Ind	Other	20	RD	1st	Engl		M/R	Yes	61		0.64	0.13	0.25
42 Mult. PA + let in Read Recovery vs. Read Recovery	3+	yes	Ind	Other	20	RD	1st	Engl		M/R	Yes	63		0.24	0.31	0.31
Hohn & Ehri, 1983													24			
43 Segment vs. No treatment	1	no	Ind	Other	2.58	Nor	K	Engl		M/R	No	16		0.77	0.2	
44 Segment + let vs. No treatment	1	yes	Ind	Other	2.58	Nor	K	Engl		M/R	No	16		1.3	0.68	
Hurford et al., 1994													99			
45 Blend & deletion + let vs. No treatment	2	yes	Ind	Comp	12	AR	1st	Engl	M-H	M/R	No	99		0.61	0.49	
Iversen & Tunmer, 1993													64			
46 Mult. PA + let in Read Recovery vs. Read Recovery	3+	yes	Ind	Other	20.88	RD	1st	Engl		M/R	Yes	64		-0.33	0.42	-0.02
Kennedy & Backman, 1993													20			
47 Mult. PA + let vs. No treatment	3+	yes	SmG	Teach	75	RD	2nd+	Engl		M/R	Yes	20		1.43	0.39	0.53
Korkman & Peltomaa, 1993													46			
48 Blend & category + let vs. Speech therapy	2	yes	SmG	Other		AR	K	Fin		NE	No	46			.60*	.67*
Kozminsky & Kozminsky, 1995													61			
49 Mult. PA vs. Visual motor integration	3+	no	Clas	Teach	21.33	Nor	K	Hebr	Lo	NE	No	61		0.24	.57*	
Lie, 1991													208			
50 Category vs. Conceptual	1	no	Clas	Teach		Nor	1st	Norw		R	No	96			0.21	0.22
51 Segment vs. Conceptual	1	no	Clas	Teach		Nor	1st	Norw		R	No	102			0.62	0.67

(continued)

APPENDIX Studies in the PA training database, their characteristics and effect sizes (continued)

	Characteristics of training					Characteristics of participants				Features of design				Effect sizes		
Author and year, treatment vs. control	Number skills	Letters	Training unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assignment	Fidelity	*N* (Case)	*N* (Study)	PA	Reading	Spelling
Lovett, Barron, Forbes, Cuksts, & Steinbach, 1994													19			
52 Segment & blend + let vs. Whole word	2	yes	Ind	Comp	18	RD	2nd+	Engl		R	No	13			1	0.02
53 Onset-rime + let vs. Whole word	1	yes	Ind	Comp	18	RD	2nd+	Engl		R	No	13			0.53	0.15
Lundberg, Frost, & Petersen, 1988													383			
54 Mult. PA vs. No treatment	3+	no	Clas	Teach	48	Nor	K	Dan	Lo	NE	No	383		0.74	0.19	.60*
McGuinness, McGuinness, & Donohue, 1995, Study 2													42			
55 Mult. PA + let in Montessori vs. No treatment	3+	yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27		0.15	1.11	
56 Mult. PA + let in whole language vs. No treatment	3+	yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27		0.37	1.22	
Murray, 1998													48			
57 Category, LS vs. Language, LS	1	no	SmG	Other	4.5	Nor	K	Engl		R	Yes	30		-0.11	0.27	
58 Segment & blend, LS vs. Language, LS	2	no	SmG	Other	4.5	Nor	K	Engl		R	Yes	30		0.41	0.07	
O'Connor & Jenkins, 1995													10			
59 Segment + let to spell vs. LS, read	1	yes	Ind	Other	3.33	AR	K	Engl		M/R	No	10		0.41	0.9	1.24
O'Connor, Jenkins, & Slocum, 1995													67			
60 Segment & blend, LS vs. LS	2	no	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45		2.69	1.64	
61 Mult. PA, LS vs. LS	3+	no	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45		2.42	0.52	
O'Connor, Notari-Syverson, & Vadasy, 1996, 1998a													80			
62 Segment & blend + let vs. No treatment	2	yes	Clas	Teach	20	Nor	K	Engl		NE	Yes	66		0.62	0.11	0.73
63 Segment & blend + let vs. No treatment	2	yes	SmG	Teach	20	AR	K	Engl		NE	Yes	14		0.03	0.99	0.97
Olofsson & Lundberg, 1983, 1985													48			
64 Mult. PA scheduled vs. Nonverbal tasks	3+	no	Clas	Teach	12.25	Nor	K	Swed		NE	Yes	38		0.7	0.28	-.07*
65 Mult. PA scheduled vs. No treatment	3+	no	Clas	Teach	12.25	Nor	K	Swed		NE	Yes	26		0.27	-0.37	0.16*
Reitsma & Wesseling, 1998													70			
66 Blend on computers vs. Vocabulary computer	1	no	Ind	Comp	4	Nor	K	Dutch		NE	No	25		0.23	.42*	-.11*
67 Blend on computers vs. No treatment	1	no	Ind	Comp	4	Nor	K	Dutch		R	No	56		0.74	.27*	.28*
Sanchez & Rueda, 1991													9			
68 Segment + let vs. Percept-motor	1	yes	SmG	Other	40	RD	2nd+	Span		R	No	9		2.19	-0.05	2.09
Schneider, Kuspert, Roth, Vise, & Marx, 1997													702			
69 Mult. PA vs. No treatment	3+	no	Clas	Teach	43.75	Nor	K	Germ		NE	No	371		0.7	0.22	.27*
70 Mult. PA vs. No treatment	3+	no	Clas	Teach	20	Nor	K	Germ		NE	Yes	331		0.82	0.05	.38*
Solity, 1996													24			
71 Segment & blend vs. Story	2	no	SmG	Other	14.75	Nor	Pre	Engl		M/R	Yes	24		0.52	1.18	
Tangel & Blachman, 1992													149			
72 Segment & category + let vs. No treatment	2	yes	SmG	Teach	13.2	Nor	K	Engl	Lo	NE	No	149		1.81	0.67	0.94
Torgesen, Morgan, & Davis, 1992													48			
73 Segment & blend, LS vs. Story, LS	2	no	SmG	Other	7	AR	K	Engl	Lo	M/R	No	31		1.87	1.22	
74 Blend, LS vs. Story, LS	1	no	SmG	Other	7	AR	K	Engl	Lo	M/R	No	32		1.82	-0.05	

(continued)

APPENDIX Studies in the PA training database, their characteristics and effect sizes (continued)

	Characteristics of training					Characteristics of participants				Features of design				Effect sizes		
Author and year, treatment vs. control	Number skills	Letters	Training unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assignment	Fidelity	N (Case)	N (Study)	PA	Reading	Spelling
Treiman & Baron, 1983													28			
75 Onset-rime vs. Repeat syllables	1	no	Ind	Other		Nor	Pre	Engl	M-H		Yes	8			0.02	
76 Onset-rime vs. Repeat syllables	1	no	Ind	Other		Nor	K	Engl	M-H		Yes	20			0.13	
Uhry & Shepherd, 1993													22			
77 Segment & blend + let vs. Text	2	yes	SmG	Comp	17.33	Nor	1st	Engl	M-H	R	No	22		1.15	1.07	0.77
Vadasy, Jenkins, Antil, Wayne, & O'Connor, 1997a													35			
78 Segment & category + let vs. No treatment	2	yes	Ind	Other	54	AR	1st	Engl	Lo	R	Yes	35		0.74	0.44	0.67
Vadasy, Jenkins, Antil, Wayne, & O'Connor, 1997b													40			
79 Segment & blend + let vs. No treatment	2	yes	Ind	Other	50	AR	1st	Engl	Lo	R	Yes	40		0.42	0.27	0.4
Vellutino & Scanlon, 1987, Experiment 2													240			
80 Mult. PA + let vs. No treatment	3+	yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		1.15	0.72	
81 Mult. PA, let, word vs. Word	3+	yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		0.74	0.5	
82 Mult. PA + let vs. No treatment	3+	yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		0.33	0.47	
83 Mult. PA, let, word vs. Word	3+	yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		1.1	0.71	
84 Mult. PA + let vs. No treatment	3+	yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		0.89	0.49	
85 Mult. PA, let, word vs. Word	3+	yes	Ind	Other	2.5	RD	2nd+	Engl		R	No	30		1.01	0.48	
86 Mult. PA + let vs. No treatment	3+	yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		-0.07	0.55	
87 Mult. PA, let, word vs. Word	3+	yes	Ind	Other	2.5	Nor	2nd+	Engl		R	No	30		0.06	0.52	
Warrick, Rubin, & Rowe-Walsh, 1993, Study II													28			
88 Segment vs. No treatment	1	no	SmG	Other	5.33	AR	K	Engl		NE	No	28		0.67	1.50*	.81*
Weiner, 1994													36			
89 Mult. PA vs. No treatment	3+	no	SmG	Other	5	AR	1st	Engl	M-H	R	No	10		0.81	0.17	
90 Mult. PA vs. No treatment	3+	no	SmG	Other	5	Nor	1st	Engl	M-H	R	No	26		0.17	-0.06	
Williams, 1980													204			
91 Segment & blend + let vs. No treatment	2	yes	Clas	Teach	62.83	RD	2nd+	Engl		NE	Yes	102		0.35	1.05	
92 Segment & blend + let vs. No treatment	2	yes	Clas	Teach	28.13	RD	2nd+	Engl		R	Yes	102		1.11	0.97	
Wilson & Frederickson, 1995													48			
93 Onset-rime + let vs. No treatment	1	yes	SmG	Other	20.67	RD	2nd+	Engl	Lo	NE	Yes	48		0.12	0.17	0.49
Wise, Ring, & Olson, 1999													122			
94 Mult. PA w/ article + let vs. Artic, LS	3+	yes	SmG	Comp	42	RD	2nd+	Engl		R	Yes	80		0.65	0.15	0.05
95 Mult. PA + let vs. Article, LS	3+	yes	SmG	Comp	42	RD	2nd+	Engl		R	Yes	85		0.66	0.28	0.3
Wise, Ring, & Olson, in press													200			
96 Mult. PA + let vs. Recip. Teach	3+	yes	Ind	Comp	24.98	RD	2nd+	Engl		R	No	200		0.77	0.23	0.05

Abbreviations:
Artic = Articulation
LS = letter-sound training provided separately
Category = categorization or identity training
Meta. = metacognitive activities to understand purposes, use of PA
Mult. = Multiple PA in 3 or more skills
Recip. Teach = Reciprocal teaching strategies learned and applied to reading
* = Effect sizes were drawn from follow-up test points.