CHAPTER

2

Alphabetics

CHAPTER 2

Alphabetics

Part I

Phonemic
Awareness
Instruction

PART I: PHONEMIC AWARENESS INSTRUCTION

Executive Summary

Introduction

When today's educators discuss the ingredients of effective programs to teach children to read, phonemic awareness (PA) receives much attention. However, not everyone is convinced. In education, particularly in the teaching of reading over the years, the choice of instructional methods has been heavily influenced by many factors, not only teachers' own frontline experiences about what works, but also politics, economics, and the popular wisdom of the day. The pendulum has swung back and forth between holistic, meaning-centered approaches and phonics approaches without much hope of resolving disagreements. Meanwhile, substantial scientific evidence has accumulated purporting to shed light on reading acquisition processes and effective instructional approaches (Anderson et al., 1985; Adams, 1990; Snow, 1998). Many studies investigating the effectiveness of phonemic awareness instruction have contributed to this body of evidence. Proponents believe that this research holds promise of placing reading instruction on a more solid footing and ending the periodic upheavals and overhauls of reading instructional practices.

The purpose of this report of the National Reading Panel (NRP) was to examine the scientific evidence relevant to the impact of phonemic awareness instruction on reading and spelling development. In the analyses conducted, the NRP sought answers to questions such as the following: 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 appropriately to yield scientifically valid findings? What does a careful analysis of the findings reveal? How applicable are these findings to classroom practice? To evaluate the adequacy and strength of the evidence, the NRP conducted a meta-analysis. The literature was searched to locate all experimental studies that included a PA treatment and a control group and that measured reading as an outcome of the treatment.

There were several reasons why phonemic awareness instruction was selected for review and analysis. 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 their first 2 years in school. This evidence suggests the potential instructional importance of teaching PA to children. Many experimental studies have evaluated the effectiveness of PA instruction in facilitating reading acquisition. Results are claimed to be positive and to provide a scientific basis documenting the efficacy of PA instruction. There is currently much interest in PA programs among teachers, principals, and publishers. State adoption committees have prescribed the inclusion of PA training in reading instruction materials approved for use in schools. It is thus important to determine whether PA instruction lives up to these claims and, if so, to identify circumstances that govern its effectiveness.

Phonemes are the smallest units constituting spoken language. English consists of about 41 phonemes. Phonemes combine to form syllables and words. A few words have only one phoneme, such as <u>a</u> or <u>oh</u>. Most words consist of a blend of phonemes, such as <u>go</u> with two phonemes, or <u>check</u> with three phonemes, or <u>stop</u> with four phonemes. Phonemes are different from graphemes, which are units of written language and which represent phonemes in the spellings of words. 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. The following tasks are commonly used to assess children's PA or 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/ /1/?” (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 are there in ship? ” (three: /š/ /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 the studies reviewed by the NRP, researchers used one or several of these tasks to assess how much PA children possessed before training and how much they had learned at the end of training. Also, these tasks were the basis for activities that children practiced during training. In some of the studies, children were taught to perform these tasks with letters, for example, segmenting words into phonemes and representing each with a grapheme. In other studies, phoneme manipulation was limited to speech.

To be clear, PA instruction is not synonymous with phonics instruction that entails teaching students how to use grapheme-phoneme correspondences to decode or spell words. PA instruction does not qualify as phonics instruction when it teaches children to manipulate phonemes in speech, but it does qualify when it teaches children to segment or blend phonemes with letters.

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. Although most English 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 to graphemes is difficult. This is because spoken language is seamless; that is, there are no breaks in speech signaling where one phoneme ends and the next one begins. Rather, phonemes are folded into each other and are coarticulated. Discovering phonemic units requires instruction to learn how the system works.

Methodology

How was the analysis of the research literature conducted?

Before conducting a meta-analysis, the NRP systematically searched the research literature relevant to PA instruction. After a methodology established by the Panel was followed, appropriate key words were entered to identify relevant studies in ERIC and PsycINFO. The search was limited to articles appearing in journals written in English, but no limit was placed on the year of publication. This yielded a total of 1,962 potentially relevant articles. Abstracts were printed and screened. In addition, references listed in these articles and in several review papers were hand-searched and screened. To qualify for analysis, studies had to meet the following criteria:

1. Studies had to adopt an experimental or quasi-experimental design with a control group or a multiple baseline method.
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 training 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.

Applying these procedures, the NRP found 52 articles from which 96 instructional comparisons were drawn. In each comparison, one group of children was taught PA while a control group received either another type of instruction or regular classroom instruction. Following training, the two groups were compared in their ability to read.

The primary statistic used in the NRP analysis was "effect size," the extent to which performance of the treatment group exceeded performance of the control group. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of PA instruction. An effect size of 0 indicates that treatment and control group means were identical, revealing that training had no effect. To judge the strength of an effect size, a value of 0.20 is considered small, 0.50 is moderate, and 0.80 is large. For each comparison, three effect sizes were calculated to determine whether PA instruction improved children's phonemic awareness, reading, and spelling.

The studies in the NRP database varied in many respects. These variations showed whether effect sizes were bigger under some conditions than others. The NRP compared effect sizes associated with the following variations:

- Type of test: a standardized test was used or a test devised by experimenters.
- Time of test: Outcomes were measured right after instruction or after a delay.
- Type of PA training: Children received instruction that focused on one type of PA or two types of PA, or they were taught three or more types of PA skills.
- Use of letters: Children were taught to manipulate phonemes using letters, or they were taught to manipulate phonemes in speech only.
- Size of groups: Children were taught individually or in small groups or in larger classroom groups.
- Trainer: The source of the instruction was the children's classroom teacher or a researcher or a computer.
- Length of instruction: Instruction varied from 1 hour to 75 hours.
- Reading level of students: The children receiving instruction were at risk for developing reading problems, or were reading disabled, or were normally developing readers.
- Grade level: The children were preschoolers, kindergartners, 1st graders, or 2nd through 6th graders.
- Socioeconomic status (SES): The children were low SES or middle-to-high SES.

In addition, the NRP examined various features of the experiments to determine whether those showing strong effects were well designed or weakly designed. Among the design features examined were whether children were randomly assigned to treatment and control groups, whether the size of the sample was small or large, and whether the study met criteria of rigor specified in a critique by Troia (1999).

Results and Discussion

What do results of the meta-analysis of PA instruction studies show?

The NRP examined whether PA instruction was significantly better than alternative forms of training in helping children acquire phonemic awareness and enabling them to apply this skill in their reading and spelling. Results were positive. The overall effect size on PA outcomes was large, 0.86. The overall effect size on reading outcomes was moderate, 0.53. The overall effect on spelling was also moderate, 0.59. Effects were significant on followup tests given several months after training ended. Effects were significant on measures of children's ability to read words and pseudowords as well as their reading comprehension. Effects were significant on standardized tests as well as experimenter-devised tests. These findings show that teaching children to manipulate phonemes in words was highly effective across all the literacy domains and outcomes. Effects of training did not generalize to performance on math tests, indicating that halo/ Hawthorne effects did not account for the findings.

What were the effects of moderators on learning phonemic awareness?

The NRP examined whether PA training was effective under more specific conditions. Children acquired PA successfully under all conditions, but some conditions produced larger effects than others. Effect sizes were larger when children received focused and explicit instruction on one or two PA skills than when they were taught a combination of three or more PA skills. Instruction that taught phoneme manipulation with letters helped normally developing readers and at-risk readers acquire PA better than PA instruction without letters. When children were taught PA in small groups, their learning was greater than when they were taught individually or in classrooms. The length of time spent teaching children was influential, with treatments lasting from 5 to 18 hours producing larger effect sizes than shorter or longer treatments. Classroom teachers were very effective in teaching PA to children. Also, computers were effective. Although all levels of readers acquired PA successfully, effect sizes were greater for children who were beginning readers at risk for reading failure and normally progressing readers than for older disabled readers. Students in the lower grades, preschool, and kindergarten, showed larger effect sizes in acquiring PA than children in 1st grade and above. Children learning to read in English showed larger effects than children learning to read in other alphabetic languages. However, SES level exerted no impact on effect size, indicating that low and mid-to-high SES children benefited similarly from PA training in acquiring phonemic awareness.

What were the effects of moderators on learning to read?

The impact of these specific conditions on the amount of transfer from PA training to other reading skills was also examined. For example, transfer was greater when experimenter-devised tests were used to measure reading skills than when standardized tests were used. This was not surprising, given that standardized tests tend to be less sensitive. Teaching that focused on one or two types of PA manipulations yielded larger effect sizes than teaching three or more PA skills. Teaching children to manipulate phonemes using letters produced bigger effects than teaching without letters. Blending and segmenting instruction exerted a significantly larger effect on reading development than did multiple-skill instruction. Small-group instruction produced larger effect sizes on reading than individual instruction or classroom instruction, albeit in an unanticipated fashion. Specifically, the longer the training program, the smaller the effect size. Significant improvement in reading skills following PA instruction was observed both in studies involving classroom teachers and in computer formats, but the degree of transfer was less than that achieved in experimentally controlled studies. Large effect sizes were obtained in studies of at-risk readers, with moderate effect sizes obtained for disabled and normally developing readers.

Moreover, preschoolers exhibited a much larger effect size on reading than did students in the other grade levels. Children learning to read in English also showed larger transfer effects to reading than children learning in other languages. The effects of PA training on reading outcomes were also influenced by SES, with mid-to-high SES associated with larger effect sizes than low SES.

What were the effects of moderators on learning to spell?

The NRP also examined how different conditions influenced the impact and transfer of PA training to spelling. The effects of PA training on spelling for disabled readers was minimal, as indicated by effect sizes that did not differ significantly from zero. This is consistent with other findings indicating that learning to spell is especially difficult for disabled readers. Because disabled readers were unevenly distributed across the conditions that were examined in relation to the effects of PA training on spelling, along with the finding of a nonsignificant effect size, data obtained from studies of disabled readers were eliminated from the database.

The effects of conditions on spelling outcomes were analyzed for at-risk and normal readers. For these groups, effect sizes involving spelling outcomes did not differ across levels of the following properties of PA training: whether one or two or multiple PA skills were taught, whether training was conducted with individuals or small groups or classroom-size groups, how long training lasted, or whether the trainer was a classroom teacher or a researcher. However, effect sizes did differ across other conditions. Teaching children to manipulate phonemes with letters exerted a much larger impact on spelling than teaching children without letters.

Also kindergartners made greater gains from PA training in spelling than 1st graders. Mid-to-high SES children showed larger effect sizes on spelling than low SES children. Children acquiring literacy in English showed larger effects on spelling than children acquiring literacy in other languages.

Did the effects of PA training arise from well-designed experiments?

The NRP examined whether significant effect sizes arose primarily from experiments with the weakest designs or whether well-designed experiments showed significant effect sizes as well. Findings indicated that rigorous designs yielded strong effects. The majority of studies used random assignment, and their effect sizes on PA and reading outcomes ranged from moderate to large. About one-third of the studies assessed trainers' fidelity to instructional procedures. Effect sizes in these studies were moderate.

Some studies compared PA treatment groups to control groups that were given another treatment, and some studies used untreated control groups. Neither type of control group consistently produced larger effect sizes, indicating that Hawthorne effects do not explain why PA training was effective. Although studies using smaller samples tended to show somewhat larger effect sizes, even those having the largest samples showed positive and significant effects that were moderate in size.

The NRP also assessed the relationship between methodological rigor and effect size by applying Troia's (1999) criteria to the studies. On PA outcomes, studies that met his criteria for the best designs produced the largest effect sizes on all five measures of rigor. On reading outcomes, effect sizes associated with the most rigorous levels were close to the largest, if not the largest, effect sizes on four out of five measures. Thus, these findings indicate that claims about the effectiveness of PA instruction are supported by evidence derived from methodologically sound studies.

Conclusions

What conclusions can be drawn from this meta-analysis of PA instruction studies?

Can phonemic awareness be taught?

Yes. The results clearly showed that PA instruction is effective in teaching children to attend to and manipulate speech sounds in words. Findings of the meta-analysis revealed not only that PA can be taught but also that PA instruction is effective under a variety of teaching conditions with a variety of learners.

Does phonemic awareness instruction assist children in learning to read? If so, which students benefit?

Yes. Results of the meta-analysis showed that teaching children to manipulate the sounds in language helps them learn to read. Across the various conditions of teaching, testing, and participant characteristics, the effect sizes were all significantly greater than chance and ranged from large to small, with the majority in the moderate range. Effects of PA training on reading lasted well beyond the end of training. PA instruction produced positive effects on both word reading and pseudoword reading, indicating that it helps children decode novel words as well as remember how to read familiar words. PA training was effective in boosting reading comprehension, although the effect size was smaller than for word reading. This was not surprising. PA instruction could be expected to benefit children's reading comprehension because of its dependence on effective word reading. However, the NRP had not expected the effect to be as strong, given that the influence is indirect. Other capabilities influence reading comprehension as well, such as children's vocabulary, their world knowledge, and their memory for text. PA instruction helped all types of children improve their reading, including normally developing readers, children at risk for future reading problems, disabled readers, preschoolers, kindergartners, 1st graders, children in 2nd through 6th grades (most of whom were disabled readers), children across various SES levels, and children learning to read in English as well as in other languages.

Does PA instruction assist children in learning to spell? If so, which students are helped?

Yes. Teaching PA was found to help children learn to spell, and its effect lasted well beyond the end of training. Some but not all types of students benefited from PA instruction. It helped kindergartners and 1st graders learn to spell. PA instruction also benefited children at risk for future reading problems and normally developing readers and was effective in boosting spelling skills in low SES as well as mid-to-high SES children. It helped children learning to spell in English as well as children learning in other languages. However, PA instruction was not effective for improving spelling in disabled readers. This is consistent with other research indicating that disabled readers have a difficult time learning to spell.

What properties of instruction make it most effective?

The NRP findings indicate that PA instruction may be most effective when children are taught to manipulate phonemes with letters, when the instruction is explicitly focused on one or two types of phoneme manipulations rather than multiple types, and when children are taught in small groups. Of course, instruction must be suited to students' level of development, with easier PA tasks appropriate for younger children. Teaching with letters is important because this helps children apply their PA skills to reading and writing. Teaching children to blend phonemes with letters helps them decode. Teaching children phonemic segmentation with letters helps them spell. If children have not yet learned letters, it is important to teach them letter shapes, names, and sounds so that they can use letters to acquire PA. PA instruction is more effective when it makes explicit how children are to apply PA skills in reading and writing tasks. PA instruction does not need to consume long periods of time to be effective. In these analyses, programs lasting less than 20 hours were more effective than longer programs. Single sessions lasted 25 minutes on average. Classroom teachers as well as computers can teach PA effectively.

Implications for Reading Instruction

Are the results ready for implementation in the classroom?

Yes. The NRP report includes many ideas that provide guidance to teachers in designing PA instruction and in evaluating existing programs. The NRP has listed references that teachers can locate for additional ideas and guidance. However, there were some important issues not addressed by the research. In implementing PA instruction in the classroom, teachers should bear in mind several serious cautions.

- Teachers should recognize that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping learners understand and use the alphabetic system to read and write. This is why it is important to include letters when teaching children to manipulate phonemes and why it is important to teach children explicitly how to apply PA skills in reading and writing tasks.
- It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness, so PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling already, whereas others may know only a few letters and have no reading skill. Nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA before beginning PA instruction. This will indicate which children need the instruction and which do not, which children need to be taught rudimentary levels of PA (e.g., segmenting initial sounds in words), and which children need more advanced levels involving segmenting or blending with letters.
- PA training does not constitute a complete reading program. Although the present meta-analysis confirms that PA is a key component that can contribute significantly to the effectiveness of

beginning reading and spelling instruction, there is obviously much more that needs to be taught to children to enable them to acquire reading and writing competence. PA instruction is intended only as a critical foundational piece. It helps children grasp how the alphabetic system works in their language and helps children read and spell words in various ways. However, literacy acquisition is a complex process for which there is no single key to success. Teaching phonemic awareness does not ensure that children will learn to read and write. Many other competencies must be taught for this to happen.

- A number of PA instructional programs were found to be effective. The studies assessing these programs are useful in identifying several factors that are important and should be considered in planning classroom instruction or in evaluating published programs that purport to teach PA. In implementing PA instruction in their classrooms, teachers need to evaluate the methods they use against measured success in their own students.
- One factor that is obviously important in any effective classroom program but has not been specifically addressed in the research literature on PA instruction is motivation of the students and of the teachers. It seems self-evident that techniques to develop children's PA in classrooms should be as relevant and exciting as possible so that the instruction engages children's interest and attention in a way that promotes optimal learning. However, research has not specifically focused on this factor. Neither has the research examined the specific techniques that are most engaging for teachers. For example, none of the studies inquired whether teachers liked the programs they were given to teach. It seems self-evident that teachers will be most effective when they are enthusiastic in their teaching and enjoy what they are doing in the classroom. In selecting ways to teach PA in their classrooms, teachers need to take account of motivational aspects of programs for themselves as well as their students.
- Results of the meta-analysis should not be overinterpreted. Although most comparisons in the analysis demonstrated significant mean effect sizes, the NRP cannot infer that every teacher of every child in the studies was successful in promoting the acquisition of PA or its transfer to reading and writing. There was considerable variation within and across individual studies. Likewise, the NRP findings should not be used to dictate any oversimplified prescriptions regarding effective PA instruction, for example, how long PA training should last (e.g., 5 to 18 hours) to be most effective. There are many factors that govern the effectiveness of instruction.
- More is not necessarily better. The NRP findings indicated that PA training was effective regardless of its length. However, effect sizes were largest when training lasted less than 20 hours. This suggests that teachers should make reasoned decisions and remain flexible about the amount of time to devote to this component of their instructional programs. Children will differ in the time they need to acquire PA. The best solution is to pretest for PA skills and adjust the amount of instruction to suit individual and class needs.
- Early PA instruction cannot guarantee later literacy success. The most reasonable conclusion from the findings of the NRP analysis is that adding well-designed PA instruction to a beginning reading program or a remedial reading program is very likely to yield significant dividends in the acquisition of reading and writing skills. Whether the benefits are lasting will likely depend on the comprehensiveness and effectiveness of the entire literacy program that is taught. Additional factors that play a significant role in children's literacy acquisition are detailed in other sections of the NRP report.

Directions for Further Research

Many experiments have been conducted to test whether phonemic awareness instruction helps children learn to read. Results have been sufficiently positive to sustain confidence that this treatment is indeed effective across a variety of child and training conditions. However, there are still some questions needing further attention from researchers.

- Research is needed to identify what teachers need to know and be able to do to teach PA effectively and to integrate this instruction with other elements of beginning reading instruction or instruction directed at older disabled readers.

- Research is needed to study whether small groups are the most effective way to teach phonemic awareness and, if so, the processes and conditions that make this approach especially effective.
- Research is needed to evaluate motivational properties of PA training programs and ways of enhancing motivation and interest if they are lacking. This includes assessing whether approaches appeal to teachers as well as students. It is important to study the factors that influence whether teachers are likely to continue using programs once they are learned.
- Research is needed to determine whether and how PA might be taught more effectively using computers so that transfer to spelling as well as reading is maximized.

PART I: PHONEMIC AWARENESS INSTRUCTION

Report

Introduction

When today's educators discuss the ingredients of effective programs to teach children to read, phonemic awareness (PA) receives much attention. However, not everyone is convinced. In education, particularly in the teaching of reading over the years, the choice of instructional method has been influenced by numerous factors, not only teachers' own frontline experiences about what works, but also politics, economics, and the popular wisdom of the day. Historically, the pendulum has swung back and forth between holistic, meaning-centered approaches and phonics approaches without much hope of resolving disagreements. Meanwhile, substantial scientific evidence has accumulated purporting to shed light on reading acquisition processes and effective instructional approaches (Anderson, Hiebert, Scott, & Wilkerson, 1985; Adams, 1990; Snow, Burns, & Griffin, 1998). Many studies investigating the effectiveness of phonemic awareness instruction have contributed to this body of evidence. Proponents believe that such research holds promise of placing reading instruction on a more solid footing and ending the periodic upheavals and overhauls.

The purpose of this report is to examine the scientific evidence supporting claims about the impact of phonemic awareness instruction on reading development. The National Reading Panel (NRP) sought answers to questions such as the following: 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? What does a careful analysis of the findings reveal? How applicable are these findings to classroom practice?

There were several reasons why the Panel 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). Such evidence suggests the potential instructional importance of PA training in the development of reading skills. Second, many experimental studies have been conducted to evaluate the effectiveness of PA training in facilitating reading acquisition. Results of these studies claim to be positive and to provide a scientific basis documenting the efficacy of PA training programs. Third, there is currently much interest in PA training programs among teachers, principals, and publishers because of claims about their effectiveness in improving children's ability to learn to read. State adoption committees such as those in Texas and California have prescribed the inclusion of PA training in reading instruction materials approved for use in schools. Thus it is important to determine whether PA training programs live up to these claims and, if so, to identify the circumstances that govern their effectiveness.

In order to evaluate the adequacy and strength of the evidence, the NRP conducted a meta-analysis. The Panel located all of the experimental studies that (1) administered PA training to students, (2) that included control groups, and (3) that measured the impact of training on reading outcomes. The Panel found 52 published studies that met the NRP criteria. The studies varied in many respects. Different types of phonemic awareness skills were taught. The participants ranged from preschoolers to 6th 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 in larger classroom groups. The meta-analytic procedure allowed the Panel to examine not only whether PA instruction exerted a significant impact on reading across all of these different conditions, but also whether these variations made any difference in the size of the impact.

Assessing and Teaching Phonemic Awareness

To understand how the Panel screened and selected studies that taught PA, it is necessary to clarify what phonemic awareness is and what it is not. 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 two phonemes, or check with three phonemes, or stop with four 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. Some of the studies reviewed taught children to use letters as aids in distinguishing the separate phonemes in speech. However, the studies the Panel accepted into the database did not go beyond this to teach conventional spelling or text writing.

PA refers to the ability to focus on and manipulate phonemes in spoken words. In the studies reviewed, researchers used the following tasks to assess children's PA or 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: /š/ /I/ /p/); and
6. Phoneme deletion, which requires recognizing what word remains when a specified phoneme is removed, for example, "What is smile without the /s/?" (mile).

One question of interest in the meta-analysis was whether teaching some forms of PA helped children learn to read better than teaching other forms.

Note that the above list does not include phoneme discrimination, which refers to the ability to recognize whether two spoken words are the same or different, for example, recognizing that tan sounds different from Dan. Phoneme discrimination is simpler than PA because it requires neither conscious awareness of phonemes nor phoneme manipulation. To qualify for analysis, studies had to teach active manipulation of phonemes, not just phoneme discrimination.

Also phoneme awareness is different from phonological awareness, which is a more encompassing term referring to various types of awareness, not only PA but also awareness of larger spoken units such as syllables and rhyming words. Tasks of phonological awareness might require students to generate words that rhyme, to segment sentences into words, to segment polysyllabic words into syllables, or to delete syllables from words (e.g., what is cowboy without cow?). Tasks that require students to manipulate spoken units larger than phonemes are simpler for beginners than tasks requiring phoneme manipulation (Liberman, Shankweiler, Fischer, & Carter, 1974). PA training in the NRP set of studies very often began by teaching children to analyze larger units. For example, Lundberg, Frost, and Petersen (1988) taught children rhyming exercises and how to break sentences into words and words into syllables before they taught children to segment initial phonemes in words. However, if the programs used to teach PA did not progress to the phonemic level, then the study was not included in the NRP data set.

In a few of the studies analyzed by the NRP, 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 consonant blend that precedes the vowel, and the rime is the vowel and following consonants, for example, j-ump, st-op, str-ong. Dividing single-syllable words into these units is easier than dividing the words in other places, for example, after the vowel (Treiman, 1985). The NRP included these studies in the set because students were essentially manipulating phonemes when the onset was a single phoneme.

Some forms of PA training in the data set qualified as phonics instruction, which involves teaching students how to use grapheme-phoneme correspondences to decode or spell words. For example, Williams' (1980) ABD program taught students to use graphemes and phonemes to blend words—which is decoding. Ehri and Wilce (1987b) taught students to use graphemes and phonemes to segment words—which is spelling. Also, Wise, King, and Olson (in press) taught both segmenting and blending with letters. What distinguished the NRP studies from the general pool of phonics training studies, however, is that instruction given to treatment students 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

As mentioned above, PA measured at the beginning of kindergarten is one of the two best predictors of how well children will learn to read. In a study by Share et al. (1984), kindergartners were assessed 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, TV watching, and many more. These researchers examined which of these measures best predicted how well the children would be reading at the end of kindergarten and at the end of 1st grade. Results showed that PA was the top predictor along with letter knowledge. PA correlated 0.66 with reading achievement scores in kindergarten and 0.62 with scores in 1st grade. Of interest in our analysis was whether PA could be shown to play a causal role in learning to read.

PA is thought to contribute in 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 linked to graphemes is difficult. This is because spoken language is seamless and there are no breaks in speech signaling where one phoneme ends and the next one begins. Rather phonemes are folded into each other 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). For these reasons, it was expected that the impact of PA training on literacy would be strongest in tasks assessing children's ability to read and spell words.

Research on word reading processes has distinguished several ways to read words (Ehri, 1991, 1994). The process of decoding words never read before involves transforming graphemes into phonemes and then blending the phonemes to form words with recognizable meanings. The PA skill centrally involved in decoding is blending. To assess decoding skill, researchers often test children's ability to read pseudowords such as blig or nef.

A second way to read unfamiliar words is by analogy to known words (Gaskins, Downer, Anderson, Cunningham, Gaskins, Schommer, & the Teachers of Benchmark School, 1988; Glushko, 1979; Goswami, 1986; Marsh, Freidman, Welch, & Desberg, 1981). A common basis for analogizing is recognizing that the rime segment of an unfamiliar word is identical to that of a familiar word, and then blending the known rime with the new onset, for example, reading brick by recognizing that -ick is contained in the known word kick. Reading by analogy is thought to require the PA skills of onset-rime segmentation and blending.

Another way to read words is from memory, sometimes called sight word reading. This requires prior experience reading the words and retaining information about them in memory. In order for individual words to be represented in memory, beginning readers are thought to form connections between graphemes and phonemes in the word. These connections bond spellings to their pronunciations in memory (Ehri, 1992; Ehri & Wilce, 1987a; Rack, Hulme, Snowberg, & Wightman, 1994; Reitsma, 1983). The PA skill thought to be important for developing word memory is being able to segment pronunciations into phonemes that link to graphemes. Formulation of this concept led to the expectation PA training would benefit children's word reading, particularly when they received practice learning to read the words.

The processes involved in writing words, either by generating approximate spellings of the words or by retrieving correct spellings from memory, require phonemic segmentation skill (Griffith, 1991). Phonemic segmentation is required for spellers to select letters to represent the phonemes. Phonemic segmentation is required to help children retain correct spellings in memory by connecting graphemes to phonemes. In the analysis it was expected that PA training would benefit children's ability to spell.

Various kinds of word reading outcomes were assessed across the studies the Panel reviewed. The simplest task given to preschoolers required them to look at a word (sat) and decide whether it says sat or mat (Byrne & Fielding-Barnsley, 1991). Studies with older children gave them lists of words to read either from standardized tests or experimenter-devised tests. Also, word learning tasks were used. For example, kindergartners first reviewed four letter-sound relations and then practiced learning to read five words over several trials, am, at, mat, sat, Sam (O'Connor, Jenkins, & Slocum, 1995). Also, pseudoword reading tasks were used in which children read nonwords such as feem, hote, cliss. Spelling tasks were included as well. Younger children were given credit for inventing phonetically plausible spellings of words while older children were scored for producing correct spellings.

Some of the studies in the NRP database measured reading comprehension as well as word reading. In order 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 text. It was expected that PA training would benefit children's reading comprehension because of its dependence on effective word reading. However, the degree of influence was expected to be less than that observed with word reading because the influence is indirect.

Design Features of Phonemic Awareness Training Studies

Many correlational studies have reported strong relationships between phonemic awareness and learning to read (for reviews, see Blachman, in press; Ehri, 1979; Stahl & Murray, 1994; Wagner & Torgesen, 1987). In correlational studies, researchers measure children's ability to manipulate phonemes and also their reading ability. Typical findings show that students who have superior phonemic awareness are better readers than students with low PA. However, such findings are insufficient to show that PA was the underlying cause enabling some students to read better than others. This is because the finding does not rule out other causal explanations for the relationship. Perhaps the correlation was observed because cause operated in the reverse direction; that is, learning to read improved students' PA. Or perhaps a third factor operated as an underlying cause boosting both PA and reading, for example, vocabulary size, memory, or general intelligence.

In order to show that PA operates as a direct cause in helping children learn to read, the NRP needed to assess evidence from experimental studies with treatment and control groups. A well-designed experiment that provides strong evidence for cause should include the following steps:

1. Pretesting should be given to students before they receive any training. Pretests verify that children have not already acquired PA and hence can profit from training. Pretest performance can be compared to posttest performance on PA, reading, and spelling tasks to evaluate gains resulting from PA training. Also, pretests indicate whether

treatment and control groups were equivalent prior to training. If not, pretests can be use to equate the groups statistically when effects of training are evaluated on outcome measures.

2. The group receiving PA training should be compared to a control group that is equivalent in all respects except for receiving the PA training. Control groups may receive another type of training involving equal time but no PA instruction, or control groups may receive no special training beyond that provided in the students' classrooms at school. The use of an alternative-treatment control group is considered preferable to a no-treatment control group because the former rules out the Hawthorne effect as the explanation for any outcome differences favoring the experimental group. The Hawthorne effect occurs when a treatment group outperforms a no-treatment control group because the treated group received special attention and as a result was more motivated to perform.

3. Random assignment should be used to place students in treatment and control groups. Random assignment makes it likely that treatment and control groups do not differ systematically in any way that would explain outcome differences following training. In other words, this step helps to establish that the treatment, rather than some other factor, was the cause of any improvement in reading outcomes.

4. Posttests should be given to students following training. Posttests to assess PA verify that training worked, that the PA-trained group made greater gains than the control group. Posttests to assess reading and spelling show that PA training transferred and improved students' reading and spelling performance.

5. Followup posttests should assess the long-term effects of PA training on students' progress in reading and spelling. Between the end of training and the followup tests, both experimental and control students receive regular instruction at school but no further specialized training in PA.

Although these features characterize a well-designed experiment, there were studies in the NRP database that lacked some of these features. Because of this, the relationship between design features and outcomes was assessed. Studies varied in whether they compared performance of the PA-trained groups to performance of treated control groups or untreated control groups. If Hawthorne effects have influenced comparisons, one would expect bigger effects when PA treatment groups are compared to untreated control groups than when compared to treated control groups. However, Bus and van Ijzendoorn (1999) in their meta-analysis reported the reverse, finding bigger effects in comparisons between PA treatment groups and control groups receiving an alternative treatment. The Panel attempted replication of their findings with the NRP data set.

The Panel also assessed whether PA training affected outcomes in three types of designs: (1) in true experiments where students were randomly assigned to treatment and control groups; (2) in quasi-experiments where students were members of pre-existing groups which were not randomly assigned to treatment and control conditions; and (3) in studies where students from treatment and control groups were matched. Although random assignment is preferable, researchers may be limited to a quasi-experimental design when they evaluate PA programs in schools where classrooms already exist or when they employ as trainers teachers who are already familiar with a program and teach it to their students. The procedure of matching children on the basis of pretest scores is done to minimize any pretreatment differences between the groups being compared. In the NRP analysis, the effects of PA training separately for the three types of studies were examined.

In a recent critique of PA training studies, Troia (1999) identified several design flaws and applied these criteria to rate PA training studies for their lack of methodological rigor. To evaluate the impact of these flaws on outcomes, the Panel examined the relationship between Troia's assessments of the PA studies and the effects reported in these studies. The purpose of this analysis was to rule out the possibility that claims about PA training effects are supported mainly by poorly designed studies.

Other Features of PA Training Studies

Studies in our data set varied in the types of students who received PA training. The NRP wanted to know whether certain types of students benefited more than other types. Studies varied in the grade level of their participants and ranged from preschool to 6th grade. Studies varied in whether their students showed any signs of having reading problems. Three types of readers were distinguished across the studies. Some focused on children at risk for developing reading difficulties in the future. These were children below 2nd grade. Being at risk was defined as having low PA or low reading in 83% of the cases. Low socioeconomic status (SES) characterized only 27% of the cases. Some studies focused on children who had already fallen behind classmates in their reading, referred to as disabled readers. These were children in 1st grade and above. The remaining studies sampled children who were judged to be making normal progress in learning to read. This judgment was based on the fact that the children were not identified as having any reading problems.

One common finding reported in many correlational studies is that children who are or will become disabled readers have poor phonemic awareness, substantially below that expected of students at their reading levels (Bradley & Bryant, 1983; Bruck, 1992; Fawcett & Nicholson, 1995). Researchers have suggested that this deficiency underlies and explains their difficulty in learning to read. In the NRP analysis, the Panel examined whether PA training was effective in teaching PA to at-risk and disabled readers and whether this improved their reading and spelling performance, thus providing evidence for a causal connection.

Studies varied in how the PA training was delivered. In some studies, researchers or their specially trained assistants taught children to manipulate phonemes. In other studies, classroom teachers were the trainers. In a few studies, training was presented primarily by computers. Because classroom teachers are the purveyors of reading instruction for most children, it is important to determine whether they can teach PA effectively. If training requires specially trained personnel, then PA instruction should not be imposed on classroom teachers. In the NRP analyses, the effects of PA training were examined separately for teachers, for computers, and for researchers.

There is substantial evidence that one-to-one tutoring is the most effective form of instruction (Bloom, 1984; Cohen, Kulik, J., & Kulik, C., 1982; Glass, Cahen, Smith, & Filby, 1982; Pinnel, Lyons, DeFord, Byrk, & Seltzer, 1994; Wasik & Slavin, 1993). However, Bus and van Ijzendoorn (1999), in their meta-analysis of PA training studies, found that teaching PA to small groups of children produced a bigger impact on outcomes than teaching students individually or in classrooms. The aim was to attempt replication of this finding with the NRP data set that included more studies than those in the previous meta-analysis.

It is common wisdom that greater time spent training students yields superior learning. However, instructional time in schools is very limited because of the many subjects and skills that must be taught. The studies in the NRP data set varied in the length of time spent teaching PA to students. To address the question of how much time might be sufficient for teaching PA, the relationship between training time and effects on learning was examined.

The NRP database included PA training studies conducted not only in English but also in other languages, such as Norwegian, Finnish, Swedish, Danish, Spanish, Hebrew, Dutch, and German. In most of these languages, the grapheme-phoneme connections are more transparent than in English. Of interest was whether PA training might exert a larger impact in English because it is harder for beginning readers to discover the graphophonemic system in English than in other languages.

Methodology

Database

An electronic search of two databases, ERIC and PsycINFO, was conducted. Six terms involving phonemic awareness 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. The search was limited to articles appearing in journals written in English, but no limit was placed on the year of publication. Using this procedure, the Panel located 637 articles through ERIC, and 1,325 articles through PsycINFO. Abstracts were printed and screened. In addition, the Panel hand-searched and screened references cited in the studies located by the electronic search and in several review papers (Apthorp, 1998; Blachman, in press; Bus & van Ijzendoorn, 1999; Stahl & Murray, 1994; Troia, 1999; Wagner, 1988).

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

1. Studies had to adopt an experimental or quasi-experimental design with a control group or a multiple baseline method.
2. Studies had to appear in a refereed journal.
3. Studies had to test the hypothesis that training in phonemic awareness improves reading performance over alternative forms of training or no training.
4. Studies had to provide training 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.

From the various lists of references, the Panel identified and located 78 articles that appeared to meet our criteria. Upon closer inspection, 26 articles did not match all criteria: 5 lacked sufficient information to determine effect size; 5 lacked an adequate control group; 12 did not assess reading as an outcome; and 4 lacked appropriate phonemic awareness training. The final set of studies meeting our criteria numbered 52 (see Appendix A).

The primary statistic used in the Panel's analysis of performance on outcome measures was effect size, indicating the extent to which performance of the treatment group exceeded performance of the control group, with the difference expressed in standard deviation units. The formula used to calculate raw effect sizes for each treatment-control comparison consisted of the mean of the treatment group minus the mean of the control group divided by a pooled standard deviation.

From the 52 studies, 96 cases comparing individual treatment and control groups were derived. Because some of the studies included more than one treatment or control group, the cases included comparisons utilizing the same group more than once. There were seven treatment groups appearing twice because they were compared to two different control groups. There were 16 control groups appearing twice because they were compared to 2 different treatment groups. There was one control group appearing three times because it was compared to three treatment groups. In sum, there were 47 independent comparisons and 49 comparisons having a group that overlapped with one or at most two other comparisons. Although this meant that effect sizes were not completely independent across cases, the Panel preferred this alternative to combining treatment and control groups within studies because it was important not to obscure important moderator variables of interest. For example, Davidson and Jenkins (1994) studied three treatment groups, one taught to blend, one taught to segment, and one taught to both to segment and blend. They compared the performance of each treatment to the same control group. The Panel wanted to retain these as separate comparisons in our analysis, so the same control group was allowed to recur in three comparisons.

A few studies in the NRP database included treatment or control groups that were not deemed appropriate for analysis. One reason was that the treatment groups provided not only phonemic awareness training but also reading or writing training that was not provided to control groups, thus confounding PA training with reading and writing training. The following describes which treatment or control groups were eliminated from the analysis and why: a treatment group given decoding training 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 teacher-trainers 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 training but also the Reading Recovery© instruction given the treatment group (Iversen & Tunmer, 1993): and a control group that did not control for all of the non-PA elements of training (Lovett et al., 1994; Vellutino & Scanlon, 1987). These treatment or control groups were not included in the database.

The studies in the NRP database were coded for many characteristics that the Panel felt were important to include as moderator variables in the meta-analyses. These characteristics are listed in Table 1 (Appendix B). Various properties of phonemic awareness training were coded. Training programs varied in whether they focused on specific PA manipulations. Single-focus studies taught blending, categorization, identity, segmention, or onset-rime only. Double-focus studies involved combinations of blending, segmenting, deletion, or categorization. Global treatments taught three or more PA skills. Programs that only taught onset-rime manipulation were coded as onset-rime training, even though the training might have involved blending and segmenting (e.g., Fox & Routh, 1976). Training varied in whether children were taught to manipulate phonemes using letters or whether attention was limited to phonemes in speech. Training that had children manipulate blank markers was coded as a nonletter treatment.

The training unit varied across studies. Students were tutored individually in some studies and in either small groups or whole classrooms in other studies. The size of the small groups varied from two to seven students. The identity of trainers varied across studies. The Panel compared classroom teachers to others who were mostly researchers or trained assistants. Credentialed teachers who conducted the training but were not the students' classroom teacher were coded as others. In a few studies, PA training was provided mainly by computers. The Panel compared this training to training provided by noncomputers (all others). The length of training varied from 1 to 75 hours. Comparisons were conducted by dividing training time into four blocks.

Characteristics of children receiving the training were coded. Children were grouped into four categories to reflect their grade levels: preschool, kindergarten, 1st grade, and 2nd through 6th grades. Also children were grouped by reading ability. At-risk children were those judged by authors of the studies to be at risk for developing reading problems. In the majority of cases (77%), this was indicated by poor performance on PA tasks. Other indicators used in a few studies were low reading, low SES, developmental or language delays, or cognitive disabilities. Only 27% of the cases were low SES, while 37% were middle-to-high SES. These children were all below 2nd grade.

Children who had already developed reading problems were coded as disabled readers. All but three cases involved children between 2nd and 6th grade levels. The three cases involved 1st graders who qualified for Reading Recovery© programs (Hatcher, Hulme, & Ellis, 1994; Iversen & Tunmer, 1993). Being reading disabled meant reading below grade level despite at least average cognitive ability in most studies. In one study, the school's definition of learning disabled was used (Williams, 1980). In one study, students were not only reading disabled but also had neurological impairment and language learning problems (Lovett et al., 1994).

Samples of children not reported as being at risk or reading disabled were coded as normally progressing readers. These studies included children selected not to have reading problems as well as children selected without regard to reading ability. The socioeconomic level of children was coded into two categories, low SES or middle-to-high SES, based on assertions by authors. The language spoken by children and used to teach PA was coded as English or non-English. Non-English languages included Dutch, Finnish, German, Hebrew, Norwegian, Spanish, and Swedish.

Some features of the methodology used in the experiments were coded. Children were assigned to treatment and control groups in one of three ways. They were randomly assigned. Or they were members of intact groups that were not randomly assigned to conditions, referred to by researchers as nonequivalent groups. In some studies two classrooms were assigned randomly, one to the treatment and one to the control condition. These cases were categorized as nonequivalent groups. In other studies, several classrooms were assigned randomly to treatment and control conditions. These cases were categorized as random assignment. The third way of assigning children to conditions involved matching children on the basis of similar test scores. Typically, members of a match are randomly assigned, one to the treatment group and one to the control group. However, in some studies, this step was not stated explicitly; so, it is impossible to be sure that random assignment was always used.

The Panel coded studies to reflect whether fidelity to treatment was checked, that is, whether researchers observed trainers to make sure they adhered to treatment procedures. In addition, comparisons were coded for the type of control group, that is, whether or not control students received a special alternative treatment or remained untreated. The number of students participating in the comparison was coded to reflect sample size. The numbers were grouped into four blocks to distinguish sample sizes ranging from small to large.

To evaluate the relationship between the methodological quality of studies and the effect sizes found, the Panel adopted the five methodological criteria applied by Troia (1999) in his critique of the internal and external validity of PA training studies. Internal validity refers to the authenticity of cause-and-effect relationships in a study, that is, whether the treatment caused the outcome observed, or whether other variables could have impacted the outcome. 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 four 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 trainer fidelity to treatment), percentage of external validity criteria met, and 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 the NRP database. The Panel applied his ratings and rankings to the 56 cases derived from these studies. The Panel did this without checking Troia's evaluations for accuracy; so, any incorrect codings of the studies arise from Troia's procedures, not from the Panel's.

One final characteristic of the NRP studies was coded and analyzed, the year of publication. Years were cast into four blocks. Other characteristics of the studies were coded as well but were not analyzed either because there was little interest or because there was an insufficient number of cases to support a meaningful analysis.

Four individuals coded the studies and entered values into the SPSS database. The reliability of moderator-variable codes was checked by comparing codes in the database to codes generated by one of the coders who re-coded 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 that were entered into the database were verified at least twice for accuracy.

There were three outcomes of primary interest: phonemic awareness, reading, and spelling performance. Some studies included multiple tasks measuring these outcomes. These measures were combined by calculating raw effect sizes (g) for individual tasks and then averaging the effect sizes across tasks. The composite measure for reading included many different types and measures of reading. For example, word reading, pseudoword reading, reading comprehension, reading speed, time to reach a criterion of learning, and miscues were included. The phonemic awareness composite included only those measures that required manipulating phoneme-size units, not larger syllabic units. The types of manipulations in the composite included segmentation, blending, reversing, deletion, identity, and categorization. The spelling composite included measures of the quality of invented spellings as well as correct spellings of words and pseudowords.

The Panel also examined more specific outcome measures that included various types of phonemic awareness, reading, spelling, and math. The specific measures are listed in Table 1. Also of interest was a comparison of effect sizes on outcomes measured immediately after training to outcomes assessing long-term learning. Delayed posttests were administered from 2 to 36 months following training.

Meta-Analysis

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

The analysis of effect sizes across studies was conducted by giving more weight to effect sizes that were based on larger samples of participants. However, the following studies administered training 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 as the value of n in the weighting procedure for these studies had the effect of underrepresenting their effect sizes. To address this problem, the Panel used n's for the unit of analysis to convert raw effect sizes (g) to corrected effect sizes (d) in each case. Then, when composite effect sizes were calculated across cases, the individual effect sizes (d) were weighted by the number of students in the sample, not by the unit of analysis, thus ensuring that no cases were underrepresented.

The DSTAT statistical package (Johnson, 1989) was employed to determine effect sizes and to test the influence of moderator variables on effect sizes. Each moderator variable had at least two levels. The Panel tested whether the mean weighted effect size (d) at each level was significantly greater than zero at $p < 0.05$, whether the individual effect sizes at each level were homogeneous ($p < 0.05$), and whether effect sizes differed significantly at different levels of the moderator variables ($p < 0.05$).

Consistency With the Methodology of the National Reading Panel

The NRP review methodology (NRP Progress Report, February 1999) was used in the search and analysis of the studies. Specifically, studies that were not published in peer-reviewed journals were excluded. All of the studies in the database employed experimental or quasi-experimental designs. The studies were coded for most of the specified categories. Categories left uncoded were those where information was rarely provided (e.g., setting [urban, rural, suburban], cost factors associated with training).

The Panel determined that a meaningful meta-analysis could be conducted on the data. The coding of moderator variables and the means and standard deviations that were used to calculate effect sizes were verified by checking all of them at least twice. Intercoder reliability was conducted on the moderator variables and agreement exceeded the prescribed level of 90%. The data analysis followed the procedures specified.

Results

Were Effect Sizes Greater Than Zero?

The statistic used to assess the effectiveness of PA training on outcome measures was effect size that measures how much the mean of the PA-trained group exceeded the mean of the control group in standard deviation units. An effect size of 1.0 indicates that the treatment group mean was one standard deviation higher than the control group mean, revealing a strong effect of training. An effect size of 0 indicates that treatment and control group means were identical, revealing that training had no effect. To judge the strength of an effect size, values suggested by Cohen (1988) are commonly used. An effect size of 0.20 is considered small; a moderate effect size is 0.50; an effect size of 0.80 or above is large.

Mean effect sizes obtained for outcome measures and levels of the moderator variables are reported in Appendix C—Table 2 for phonemic awareness, Table 3 for reading, and Table 4 for spelling. Effect sizes were tested statistically to determine whether each was significantly greater than zero, indicating that superior performance of PA trained groups over control groups was not likely a result of chance at $p < 0.05$. Inspection across Tables 2 and 3 in Appendix C reveals that all of the effect sizes involving phonemic awareness and reading outcomes were significantly greater than zero. This indicates that training was effective in teaching phonemic awareness and in facilitating transfer to reading across all of the conditions and characteristics considered.

Inspection of spelling outcomes in Table 4 reveals that all but three effect sizes were significantly greater than zero. This indicates that, across most of the conditions and characteristics considered, phonemic awareness training transferred and improved spelling skills more than alternative forms of training or no training. Effect sizes for spelling outcomes were insignificant when computers were used in the training, and when the students trained were disabled readers or children in 2nd grade and above. As documented below, the absence of significant effects on spelling outcomes in the latter cases arose primarily because disabled readers' spelling benefited little from PA training, and these readers were overrepresented in these categories (i.e., 2nd through 6th graders, receiving PA instruction on computers).

Some of the studies evaluated the effects of PA training on an outcome not expected to be affected (e.g., mathematics). Tests to assess math were administered following training in 12 comparisons and following some delay in three comparisons. Results in Table 3 show that the effect size was nonsignificant and close to zero ($\underline{d} = 0.03$). This indicates that the effects of PA training did not influence all outcomes but rather were limited to outcomes related to literacy. These findings argue against the operation of any halo/Hawthorne effect explaining the positive effect sizes.

In sum, these findings led the Panel to conclude with much confidence that phonemic awareness training is more effective than alternative forms of training or no training in helping children acquire phonemic awareness and in facilitating transfer of PA skills to reading and spelling. PA training improves children's reading performance in various types of tasks, including word reading, pseudoword reading, and reading comprehension. Benefits are evident on standardized tests as well as experimenter-designed tests of reading and spelling. Improvement in reading and spelling is not short-lived but lasts beyond the immediate training period.

PA training improves reading performance in preschoolers and elementary students, and in normally progressing children, as well as in older disabled readers and younger children at risk for reading difficulties. PA training improves spelling performance in kindergartners, 1st graders, and at-risk students, but not in older disabled readers. PA training boosts reading and spelling in both English and non-English languages, and among low SES as well as middle-to-high SES children. Many types of PA training programs are effective for improving reading and spelling, including those that teach one or multiple types of phonemic awareness, those that incorporate letters into training, and those that limit phoneme manipulation to speech. Not only researchers but also classroom teachers and computers can deliver PA instruction effectively. Instruction can be conducted successfully with individuals as well as small groups and whole classrooms. Training does not have to be lengthy to be effective.

Were Effect Sizes Homogeneous?

In addition to determining whether mean effect sizes were significant, the Panel also tested whether the set of effect sizes was sufficiently homogeneous to render the mean effect size representative of that set. 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). The 95% confidence intervals for effect sizes presented in Tables 2 to 4 reveal how variable they were. When the pool of effect sizes is not homogeneous, the next step is to examine whether moderator variables reduce the variability among effect sizes to create homogeneity, indicating their power to explain the variance.

At the top of Tables 2, 3, and 4 in Appendix C, it is apparent that on the immediate outcome measures of PA, reading, and spelling, effect sizes were not homogeneous, as indicated by "No" in the homogeneity column. Effect sizes involving followup measures of PA and spelling outcomes were homogeneous, but followup reading effect sizes were not. Thus, there is reason to examine moderator variables that may explain effects on immediate outcomes and on followup tests involving reading outcomes.

Did Moderator Variables Influence Effect Sizes?

Studies varied in many respects as indicated in Table 1 (Appendix B). The Panel examined whether these moderator variables enhanced or limited the effectiveness of PA training for teaching PA and for facilitating transfer to reading and spelling. It is important to recognize the limitations of this type of analysis and the tentative nature of any conclusions that are drawn. Findings involving the impact of moderator variables on effect sizes cannot support strong claims about causality. Moderator findings are no more than correlational. The biggest source of uncertainty is whether there is a hidden variable that is confounded with the variable in focus and is the true cause of the difference; thus, the conclusions drawn should be regarded as tentative and suggestive rather than the final word.

Another caution to keep in mind in interpreting findings involving moderator variables is that the same 96 cases in the database do not contribute to the calculation of all effect sizes. Rather the set of cases changes across moderator variables, either because some of the studies lacked the information to be coded, they did not assess the outcome in interest, or they did not include a measure of the outcome at that test point. Any instability in the pattern of findings may arise from this source, particularly when only a few cases contribute.

Outcome Measures

The immediate goal of phonemic awareness training across these studies was to improve children's phonemic awareness. From Table 2, it is apparent that the effect size after training was large (d = 0.86), and it did not decline significantly at the followup test (d = 0.73). Thus, PA training taught phonemic awareness very effectively, and students retained their skill after training ended. Comparison of specific PA skills acquired during training indicated that effects were larger for segmentation and deletion outcomes than for blending. Perhaps blending was harder to teach, or perhaps it was easier for controls to pick up without instruction.

The strong gains in PA were observed to transfer to reading and spelling, and effects persisted through the second followup test. As evident in Table 3, reading-outcome effect sizes were moderate, and the effect size after training (d = 0.53) was equivalent to that at the first followup test (d = 0.45). A significant effect size was still present but significantly smaller at the second followup test (d = 0.23). Table 4 shows that spelling outcomes were boosted by PA training. The effect size following training (d = 0.59) was moderate and significantly greater than the effect sizes at the two delayed posttests (d = 0.37 and 0.20) that did not differ. PA training benefited children's reading and spelling performance not only on experimenter-devised (E) tests but also on standardized (S) tests, although the effect size was significantly larger with experimenter tests (d = 0.61 E vs. 0.33 S for reading; d = 0.75 E vs. 0.41 S for spelling). This is perhaps not surprising. Standardized tests are designed to assess reading and spelling across a wide range of ability levels and hence are less sensitive to differences at any one level in the range. Also, experimenter tests may be more sensitive because often they are tailored to detect the phonemes and graphemes that were taught.

Some studies assessed reading performance with pseudowords in order to measure children's ability to decode unfamiliar words. From Table 3, it is apparent that PA training benefited decoding skill. Effects were moderate and equivalent on both experimenter-devised tests (d = 0.56) and standardized tests (d = 0.49).

The effect of PA training on reading comprehension was assessed in 18 cases. From Table 3, it is apparent that training boosted reading comprehension significantly (d = 0.32), although the effect size was smaller than for word reading. This is not surprising. PA training would be expected to influence comprehension primarily through its impact on word reading. The task of reading, understanding, and remembering information in the text involves multiple processes. Not only must students read the words, but also they must do so rapidly and accurately and must construct meaning across the words and sentences. These other processing demands could be expected to dilute the influence of PA training.

Properties of PA Training

Studies varied in whether one skill, two skills, or multiple skills were taught. These skills consisted mainly of teaching children to identify or categorize phonemes, or to blend, segment, or delete phonemes, or to manipulate onset-rime units. From PA outcomes in Table 2, it is apparent that focusing instruction on one or two skills was significantly more effective for teaching phonemic awareness than focusing on multiple skills (d = 1.16 for one vs. d = 1.03 for two vs. d = 0.70 for multiple). One explanation for lower effect sizes is that children who were taught many different ways to manipulate phonemes may have become confused about which manipulation to apply when the various kinds of PA were assessed after training. Another possibility is that insufficient time was spent on any one type of PA to teach it well in the multiple condition. A third possibility is that multiple skills instruction involved teaching higher level PA skills mainly to older children having difficulty acquiring PA.

The Panel examined whether focused training in PA produced greater transfer to reading than multiple-skill training. From reading outcomes in Table 3, it is apparent that transfer was twice as great when PA training focused on one (d = 0.71) or two (d = 0.79) PA skills than when a multitude of skills were taught (d = 0.27). The advantage of focused over multiple-skill training for reading persisted at the followup test, especially for the two-skill focus that produced significantly larger effects than the one-skill focus. This indicates that teaching two PA skills to children has greater long-term benefit for reading than teaching only one PA skill or teaching a global array of skills.

As evident in Table 4, spelling effect sizes for focused and multiple skills instruction showed the same pattern. In fact, effects for the one-skill condition (d = 0.74) and the two-skill condition (d = 0.87) were over three times as large as the effect size for the multiple condition (d = 0.23). These findings suggest that focused PA instruction may benefit spelling more than multiple skill instruction does. However, it is likely that the lower effect size in the multiple condition arose because disabled readers dominated this category and PA instruction did not improve their spelling (see below).

Various types of phoneme manipulations might be taught. However, two types, blending and segmenting, are thought to be directly involved in reading and spelling processes. Blending phonemes helps children to decode unfamiliar words. Segmenting words into phonemes helps children to spell unfamiliar words and also to retain spellings in memory. A number of studies examined PA training that taught children to blend and segment phonemes. To assess its value, the Panel compared the effect size for this treatment to the effect size for the multiple (3 or more skills) treatment. As evident in Table 2 reporting PA outcomes, neither form was more effective than the other for teaching PA. However, as evident in Table 3 for reading outcomes, teaching students to blend and segment benefited their reading much more (d = 0.67) than did a multiple-skills approach (d = 27). As shown in Table 4, the blending and segmenting treatment also produced a larger effect on spelling performance (d = 0.79) than did the multiple skill treatment (d = 0.23), but very likely this resulted from disabled readers' dominating the multiple treatment condition (see below). From these findings, the Panel concludes that blend-and-segment training benefited children's reading more than multiple skills training did.

Also of interest was whether some types of single phoneme manipulation activities, for example, blending, segmenting, or categorizing, were more effective than other types. However, in examining the database, there were too few instances of each type to permit comparison; so, this question was not addressed in the Panel's analysis.

Studies in the database differed in whether or not children were taught to manipulate phonemes using letters during training. For example, some children learned to segment words into phonemes by selecting plastic letters for the sounds they spoke, whereas other children only spoke the sounds or they represented the sounds with unmarked tokens. Of interest was whether letters might improve children's learning because they provide concrete, lasting symbols for sounds that are short-lived and hard to grasp. From PA outcomes in Table 2, it is apparent that children trained with letters did not acquire stronger PA (d = 0.89) than children trained without letters (d = 0.82). The absence of a difference may have occurred, however, because almost all comparisons involving disabled readers fell in the letter use category, and disabled readers exhibited smaller effect sizes than nondisabled readers on PA outcomes (see Table 2). As described below, when effects of letter use were examined after disabled readers were removed from the database, a significant advantage of letter use was detected. From these findings, the Panel concludes that teaching PA with letters is more effective in helping nondisabled readers acquire phonemic awareness than teaching PA without letters.

It was expected that teaching PA with letters would facilitate greater transfer to reading and spelling than teaching PA without letters. This is because reading and spelling processes require knowing how phonemes are linked to letters. From reading outcomes in Table 3, it can be seen that teaching children to manipulate phonemes with letters created effect sizes almost twice as large as teaching children without letters (d = 0.67

vs. 0.38). The same pattern persisted at the followup test as well (d = 0.59 vs.0.36). Likewise, letters benefited spelling more than no letters, with the effect size almost twice as great (d = 0.61 vs. 0.34). These findings reveal that PA training makes a stronger contribution to reading and spelling performance when the training includes teaching children to manipulate phonemes with letters than when training is limited to speech.

Studies varied in whether PA training was provided to individual students or small groups or classrooms of students. From PA outcomes in Table 2, it is evident that the most effective way to teach PA was in small groups. The effect size produced by small groups was very large (d = 1.38), over twice the size of effects for individuals (d = 0.60) and classrooms (d = 0.67). This was surprising given that it is easier to tailor instruction and corrective feedback when students are taught individually, and it was expected that this advantage would make individual instruction more effective. Explanations for the effectiveness of training in groups promoting the acquisition of PA may involve enhanced attention, social motivation to achieve, or observational learning opportunities.

The superior PA skills acquired by children taught in small groups transferred and boosted their reading and spelling performance as well. Effect sizes on reading outcomes for small groups were d = 0.81 on the immediate posttest and d = 0.83 on the followup posttest. In contrast, effect sizes for children taught individually or in classrooms ranged from d = 0.30 to 0.45 on the immediate and delayed posttests. On spelling outcomes, small group instruction produced a larger effect size than individual instruction did, but the small group effect size did not differ from the classroom effect size (see Table 4).

The possibility that small group effect sizes might be inflated for statistical reasons was considered. Studies that treated groups as the unit of analysis in statistical comparisons may have exhibited larger effect sizes than studies using individuals as the unit of analysis because the standard deviations of group means are smaller than the standard deviations of individual scores. However, there were only five studies that used groups as the statistical unit of analysis, and these contributed only seven cases (15%) to the total of 45 cases in which children were trained in small groups. The small number of instances serves to rule out this explanation for the larger effect sizes associated with small group training.

The length of time allocated for PA training varied from 1 hour to 75 hours across studies. Cases were grouped into four time blocks to determine whether there was an optimum length of time for teaching PA. From phonemic awareness outcomes in Table 2, it is evident that effect sizes were significantly 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 less effective for teaching PA, in fact, only half as effective (d = 0.61 and 0.65).

On reading outcomes, training programs that were long-lasting yielded a significantly smaller effect size than shorter training programs as shown in Table 3. Effect sizes for the three shorter time blocks did not differ. The same pattern was evident on spelling outcomes.

These findings run counter to the expectation that more extensive training in PA should enable children to acquire superior phonemic awareness with stronger benefits for reading and spelling. These findings suggest that PA training does not need to be lengthy to exert its strongest effect on reading and spelling. However, caution is needed in drawing conclusions. There are various reasons why effect sizes might have been smaller when training was extensive. Perhaps the goals of instruction were more complex and harder to achieve. Or perhaps the students who received extended training were harder to teach. Alternatively, perhaps shorter instruction is better. The value of PA instruction may be to initiate insight into the alphabetic system. Adding further nuances or complexities may erode learning by producing confusion or boredom. In sum, the optimum length of PA training remains an issue needing further research.

Classroom teachers are the primary purveyors of reading instruction so, it is important to verify that they can teach PA effectively. Results of the analysis of phonemic awareness outcomes (see Table 2) showed that the effect size produced by classroom teachers was large (d = 0.78) although not as large statistically as that produced by others, consisting mainly of researchers (d = 0.94). This is not surprising, given that researchers were the ones who devised the training procedures in all of the studies.

PA training delivered by teachers transferred to reading and spelling. In the case of reading outcomes, the effect size associated with classroom teachers was significantly smaller (d = 0.41) than the effect size of researchers (d = 0.64). Of course, in these studies, neither teachers nor researchers intervened and helped 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 effect sizes associated with classroom teachers were significantly greater than effect sizes associated with 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 (see below).

There were only seven studies that used computers to teach PA. Ten treatment-control comparisons were derived from these studies. From PA outcomes in Table 2, it is apparent that computers produced a moderately strong effect size on the acquisition of PA (d = 0.66) although it was significantly less than the effect size for other forms of instruction (d = 0.89). The phonemic awareness that children learned from computers transferred and improved their reading performance on the immediate posttest (d = 0.33), but computers did not improve reading as much as other forms of PA instruction (d = 0.55). In contrast to the effects on reading, computer instruction exerted no significant effect on spelling outcomes (d = 0.09). One reason is that most of the computer comparisons involved disabled readers whose spelling performance did not benefit from PA training. From these findings the Panel concludes that computers are effective for teaching PA and for promoting transfer to reading, but they may be ineffective for teaching spelling to disabled readers.

Characteristics of Students

Some of the studies in the database targeted younger students at risk for future reading problems and older students classified as disabled readers. Both groups have been found to exhibit excessive difficulty manipulating phonemes in words (Bradley & Bryant, 1983; Juel, Griffith, & Gough, 1986; Juel, 1988). PA training programs were designed to remediate these readers' PA problems. Three types of readers were coded in the database: at-risk, disabled, and normally progressing readers. A comparison of phonemic awareness outcomes across the three groups revealed that although effect sizes were moderate to large in all cases, they were signficantly smaller for disabled readers (d = 0.62) than for at-risk (d = 0.95) and normally progressing readers (d = 0.93). This suggests that it was harder to improve PA in reading disabled students than in nondisabled students, perhaps because the disabled readers were older and relatively more advanced in PA skills with less room for gains than the younger beginning-level readers. Also it was the case that disabled readers were taught more advanced forms of PA (i.e., segmenting and blending with letters) than the younger students. At-risk readers were found to gain as much from PA training as normally developing readers. This indicates that having low PA when training began did not hinder at-risk readers in acquiring PA.

One might expect this pattern to be replicated on reading outcomes. However, Table 3 reveals that at-risk children showed bigger transfer effects in their reading (d = 0.86) than normal and disabled students whose effect sizes were equivalent (d = 0.47 for normals and d = 0.45 for disabled). Effect sizes on followup reading tests showed the same pattern except that the effect size for at-risk students was even larger (d = 1.33), while the effect sizes of the other two groups were smaller (d = 0.30 for normals and 0.28 for disabled). These findings indicate that PA training gives at-risk students a bigger boost in reading than it gives normals or disabled readers.

The effect of PA training on spelling outcomes differed among the three reader groups. Effect sizes were large and similar for at-risk (d = 0.76) and normal readers (d = 0.88). However, as indicated above, the effect size was much smaller, in fact, not significantly different from zero for disabled readers (d = 0.15). These findings show that PA training is not effective for improving disabled readers' spelling skills, perhaps because their spelling skills are much harder to remediate than their reading skills. In contrast, PA training was found to transfer to spelling in at-risk and normally progressing readers, indicating that PA training does benefit spelling in nondisabled readers.

The Panel also examined the effects of PA training at various grade levels: preschool, kindergarten, 1st grade, and 2nd through 6th grades. From PA outcomes in Table 2, it is evident that preschoolers showed a very large effect size in acquiring PA (d = 2.37). However, only two cases contributed to this value, making it less reliable. The effect on PA outcomes in kindergarten (d = 0.95) was significantly larger than the effect in 1st grade (d = 0.48) and in 2nd through 6th grades (d = 0.70). The latter two effect sizes did not differ. These findings indicate that younger students gained the most PA, not surprisingly since they started out with the least PA.

Effect sizes for reading outcomes in Table 3 reveal that PA training transferred to reading to a similar extent for kindergartners, 1st graders, and 2nd through 6th graders (ds from 0.48 to 0.49). The effect size for preschoolers was much larger (d = 1.25). The same pattern was not apparent on spelling outcomes, as evident in Table 4. Transfer of PA training to spelling was greater among kindergartners (d = 0.97) than among 1st graders (d = 0.52). There was no transfer to spelling among the 2nd through 6th graders for whom the effect size did not differ from zero (d = 0.14). (Spelling was not measured in the preschool studies.) The absence of an effect on spelling among the older children arose primarily because the majority of the cases in 2nd through 6th grades (78%) consisted of disabled readers who failed to show transfer effects from PA training to spelling (see below).

The Panel examined the relationship between the socioeconomic status of students across studies and the size of effects produced by PA training. As evident for PA outcomes in Table 2, low and mid-to-high SES levels did not differ, and both levels showed large effect sizes in acquiring PA. However, transfer to reading and spelling was significantly greater among among mid-to-high SES 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, disabled reader effect sizes did not contribute to SES effect size calculations.

The NRP database included many studies conducted in English-speaking countries as well as a smaller number of studies conducted in countries speaking languages other than English. A comparison of effect sizes revealed that PA training exerted a larger impact on the acquisition of PA by English-speaking students (d = 0.99) than by the non-English students (d = 0.65). Transfer to reading outcomes was also greater for English students (d = 0.63) than for others (d = 0.36) on the immediate test but not the followup test. However, there were no differences in effects sizes on spelling outcomes.

A possible reason for the absence of effects on spelling is that most of the studies involving disabled readers were in the pool of English studies. This may have suppressed the English effect size in spelling. To check on this, effect sizes were recalculated with the reading-disabled (RD) comparisons removed (see below). Results confirmed suspicion; they changed from no effect on spelling to a significant effect favoring English (d = 0.95) over non-English (d = 0.51).

One intriguing 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 other languages; so, explicit training may make a bigger contribution to clarifying phoneme units and how they link to graphemes in words for English-speaking students.

Analysis of Moderator Effects With Disabled Readers Removed From the Database

In the analysis of effects associated with the three types of readers, effect sizes were significantly smaller for disabled readers than for at-risk and normal readers on two outcomes, phonemic awareness and spelling. In fact, on the spelling outcome, no significant effect of PA training was detected for disabled readers. Moreover, the pool of spelling effect sizes for disabled readers was homogeneous, indicating that no further analysis of moderator variables was needed to locate cause and allowing us to conclude that PA training does not improve spelling in disabled readers.

In the NRP database, there were 17 comparisons involving disabled readers (18% of the total comparisons). The Panel worried that conclusions about how moderator variables regulate the impact of PA training on phonemic awareness and spelling outcomes might be different if cases involving disabled readers were removed from the database. As discussed above, in our analysis of English and non-English studies, findings changed for spelling outcomes with reading disabled cases eliminated. This was because the distribution of disabled reader cases was uneven, with most cases falling in the English pool of effect sizes. There were other moderator variables with an uneven distribution of disabled readers across levels as well. Disabled readers were older (mostly in grades 2 through 6), they tended to receive PA 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 teachers were most often the trainers.

To examine whether findings involving these moderators would be different without disabled readers, effect sizes were re-analyzed after removing disabled reader comparisons from the database. The following specific moderator variables were re-analyzed: PA skills taught, use of letters, grade, language, training unit, teachers vs. others as trainers, and length of training. Computer effects were not re-analyzed because there were too few cases.

Findings involving spelling outcomes were altered for several moderators when disabled readers were removed. Findings involving PA outcomes were altered for one moderator. However, findings were not altered at all in the analyses of reading outcomes. Results are given in Table 5 (Appendix D).

Comparison of the number of cases contributing effect sizes to spelling outcomes with and without disabled readers (Tables 4 vs. Table 5) reveals that the numbers dropped substantially in the following categories: three or more PA skills taught (drop from ten to three cases), letters manipulated (from 27 to 17 cases), individual instruction (from 14 to 8 cases), small group instruction (from 20 to 15 cases), training lasting 20 to 75 hours (from 18 to 9 cases), researcher as trainer (from 30 to 20 cases), 2nd through 6th graders (from 8 to 0 cases), English language (from 32 to 22 cases). The same comparison for PA outcomes (Table 2 vs. Table 5) reveals that in the category of letters manipulated, the number dropped from 39 to 25 cases. Declines in the other categories listed in Table 5 were minimal. This verifies that disabled readers were unevenly distributed across levels of these moderators. The SES variable was not affected and hence not re-analyzed because most studies involving disabled readers did not report the SES level of the readers.

In all but one analysis of spelling outcomes, the pattern of effect sizes changed when disabled readers were removed from the database. PA teaching that focused on one or two skills was no longer superior to multiple PA skill teaching. (However, note in Table 5 that there were only three cases left in the multiple skills category, raising doubt about the reliability of this effect size.) Small group instruction no longer produced better transfer to spelling than individual instruction. Training periods lasting 20 or more hours were no longer less effective than shorter training periods. Classroom teachers no longer differed from researchers in facilitating transfer to spelling. In the analysis of spelling outcomes across grades, the 2nd through 6th grade category had no comparisons to contribute effect sizes. The loss of cases in the upper grades shows that disabled readers clearly dominated effect sizes in this category. The greater effect of PA training on spelling among kindergartners than 1st graders remained the same.

There were two moderators that did not differentially influence spelling or PA outcomes when the whole database was analyzed; but when disabled reader effects were removed, significant differences appeared. As evident in Table 5, language now impacted spelling effect sizes, with English-speaking students benefiting more from PA training than non-English-speaking students. Also, letter use now impacted phonemic awareness effect sizes such that children who manipulated letters acquired more PA than children who did not. Removal of disabled readers rendered findings for these moderators consistent across all three outcomes. That is, language exerted the same impact on PA, reading, and spelling outcomes, with English producing larger effects than non-English. Also letter use exerted the same impact on PA, reading and spelling, with letter manipulation producing larger effects than no letters.

In sum, these findings support the following conclusions. PA training does not improve spelling in disabled readers, but it does improve spelling in normally developing readers below 2nd grade and children at risk for future reading problems. Among nondisabled readers, the benefit to spelling is positive and does not depend on whether one or two or multiple PA skills are taught, whether instruction is delivered to individuals or to small groups, how long training lasts, or whether teachers or researchers are the trainers. However, the benefit to spelling among nondisabled readers does depend upon the language, with PA training in English exerting a bigger impact on spelling than PA training in other languages.

Regarding the acquisition of phonemic awareness by nondisabled readers, our findings support the conclusion that PA training is more effective when it is taught by having children manipulate letters than when manipulation is limited to speech.

It is important to note that the pattern of effect sizes on reading outcomes remained unchanged when comparisons involving reading disabled students were removed. Specifically, teaching one or two PA skills still resulted in larger effect sizes on reading than teaching a multitude of PA skills. Small groups still produced superior transfer to reading than individual instruction. Lengthy training periods still yielded smaller effects on reading than shorter training periods. These findings serve to sustain our conclusions about the influence of moderators on reading outcomes.

Design Features

Studies in the database varied in methodological rigor. The Panel examined some of these properties to see whether design weaknesses inflated effect sizes.

Studies varied in whether or not subjects were randomly assigned to treatment and control groups. In some cases, nonrandom, nonequivalent groups were assigned to treatment and control conditions. In some cases, group assignment involved matching individual children on the basis of similar test scores. Effect sizes for the three assignment types were determined (see Tables 2, 3, and 4 in Appendix C). Comparison of PA outcomes revealed very similar effect sizes that did not differ statistically and ranged from 0.83 to 0.92. Comparison of reading outcomes revealed that the effect size for randomly assigned groups ($\underline{d} = 0.63$) was significantly greater than the effect size for nonequivalent groups ($\underline{d} = 0.40$). However, the opposite was found on spelling outcomes, with nonequivalent groups showing a significantly larger effect size ($\underline{d} = 0.86$) than random groups ($\underline{d} = 0.37$). These findings show that larger effect sizes in our database did not consistently arise from weaker designs involving nonequivalent groups. Moreover, average effect sizes for the most rigorous assignment procedure, random assignment, ranged from low-moderate to large.

Some researchers in the database administered fidelity checks to ensure that trainers adhered to prescribed training procedures, whereas other researchers did not, or at least did not report, doing this. A comparison revealed that significantly larger effect sizes arose in studies not checking for fidelity than in studies checking for fidelity. This was true across all three outcome measures (see Tables 2, 3, and 4 in Appendix C). Although weaker studies involving lack of fidelity checking were associated with larger effects, fidelity studies nevertheless yielded significant effects that were moderate in size. This verifies that lack of rigor in fidelity checking does not explain effect sizes in the NRP database.

Bus and van Ijzendoorn (1999) reported an unexpected finding in their PA meta-analysis, that studies using treated control groups yielded larger effect sizes than studies using untreated control groups. This finding was examined in the present meta-analysis. Results were mixed. On PA outcomes, the two types of control groups did not yield significantly different effect sizes. On reading outcomes, they did, with studies using treated controls showing larger effects than those using untreated controls, consistent with Bus and van Ijzendoorn's finding. On spelling outcomes, studies with untreated controls showed larger effects than studies with treated controls, the reverse pattern.

The foregoing results emerged from an analysis of all the studies. However, these studies varied in many respects besides the type of control group they used. In the NRP database, there were eight studies that compared PA training to both a treated control group and an untreated control group. In limiting the analysis to these studies, the Panel found that, out of 20 comparisons, ten showed bigger effects in cases using treated controls and ten showed bigger effects in cases using untreated controls across the three outcome measures. Thus, the picture arising from this analysis was mixed.

Although the findings reveal no clear pattern favoring treated or untreated control groups, the fact that studies using untreated controls did *not* uniformly yield larger effect sizes serves to challenge the commonly held belief that untreated control groups always yield larger effects. It is not the case that Hawthorne effects always prevail. Other factors appear to influence outcomes as well. Perhaps Hawthorne effects are more characteristic of older participants with better developed metacognitive sensitivities.

Among studies in the NRP database, samples included as few as nine students or as many as 383 students. To examine whether effects differed as a function of sample size, the studies were divided into blocks of approximately equal numbers of cases. Outcomes reported in Tables 2 to 4 reveal that 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 significantly greater than zero even in the largest samples shows that the PA training effects observed did not arise primarily from the weaker studies with small samples.

Recently Troia (1999) published a critique of phonemic awareness training studies. He identified several criteria to assess methodological rigor and applied these criteria to 39 PA training studies of which 29 were in the NRP database. (The remaining studies did not assess reading as an outcome so were not among the studies considered.) The Panel incorporated his summary ratings into the NRP database and examined the relationship between these evaluations and effect sizes. 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. Also he ranked the studies to indicate their overall methodological rigor. The Panel's purpose was to consider and rule out the possibility that effects of PA training were limited primarily to studies that were the least rigorous. Comparisons were grouped into blocks of three or four in order to reveal effect sizes at the various levels of rigor.

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

In Table 6, reporting effects of PA training on PA outcomes, it is apparent that across all five measures the largest effect sizes occurred for the blocks reflecting the most rigor. This shows 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. The effect size associated with the most rigorous level is close to the strongest, if not the strongest, effect size on four of the five measures: the two internal validity measures, the external validity critical flaws measure, and the overall rigor ranking. On the remaining measure, percent of external validity criteria met, the effect size is moderately strong though less so than the largest effect size. This evidence indicates that the better designed studies tended to produce stronger transfer effects in reading than the weaker studies.

In sum, although Troia (1999) finds fault with PA training studies, his findings do not undermine claims about the effectiveness of PA training for helping children learn to read. Troia's concluding plea, that researchers maintain high standards in designing their studies, is supported by Panel 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.

One final characteristic of studies examined was the year of publication. From Tables 2 and 3, it is apparent that there was one period in which a spate of PA training studies was published, from 1991 to 1994. Over twice as many studies were published during this period as during the other periods. The 1991 to 1994 studies also tended to yield larger effect sizes on PA and reading outcomes than studies in time periods before or after this. Why this occurred is not clear.

Discussion

Summary of Findings

To summarize results of the meta-analyses, the Panel examined 96 cases, each comparing a treatment group that received PA training, to a control group that received an alternative form of instruction or no special instruction; they examined effects on three main outcome variables, PA, reading, and spelling.

PA training was found to be very effective in teaching phonemic awareness to students. Effect sizes were large immediately after training (d = 0.86), and they remained strong over the long term (d = 0.73). PA training succeeded in teaching children various ways to manipulate phonemes, including segmentation, blending, and deletion. PA training was effective in teaching PA skills across all levels of the moderator variables examined.

PA training improved children's ability to read and spell in both the short and the long term. The effect size was moderate following training on reading (d = 0.53) and on spelling (d = 0.59). Tests of word reading, pseudoword reading, and reading comprehension all yielded statistically significant effect sizes on both experimenter-devised tests as well as standardized tests. Few instances occurred in which moderator variables reduced effect sizes to chance levels, and these were limited to spelling outcomes. Whereas PA training exerted strong effects on reading and spelling, it did not impact children's performance on math tests. This indicates that halo/Hawthorne effects did not explain findings and that training effects were limited to the targeted domain.

Several moderator variables were found to influence children's acquisition of phonemic awareness. PA training programs varied in whether children were taught to manipulate phonemes in one, two, or multiple ways, and in the type of phoneme manipulations taught, segmenting, blending, deleting, identifying, or categorizing phonemes, or manipulating onsets and rimes. Properties of the training procedures exerted an impact. Programs that focused on teaching one or two PA skills yielded larger effects on PA learning than programs teaching three or more of these manipulations. Instruction that taught phoneme manipulation with letters helped children acquire PA skills better than instruction without letters. Facilitation from letters was observed among at-risk readers and normally developing readers below 2nd grade. It was not possible to assess the contribution of letters among disabled readers because most studies used letters to teach PA to disabled readers.

Teaching children in small groups produced larger effect sizes on PA acquisition than teaching children individually or in classroom-size groups. Classroom teachers produced large effect sizes, indicating that they were very successful in teaching PA to students, although researchers produced somewhat larger effects. Computers also taught PA effectively. The length of training influenced PA acquisition. Effect sizes were larger when PA instruction lasted from 5 to 18 hours than when either less or more time than this was spent.

Characteristics of students influenced how much phonemic awareness they acquired from training. Disabled readers showed smaller effect sizes than at-risk students or normally progressing readers, indicating that PA was harder for disabled readers to learn. Also students in the lower grades, namely preschool and kindergarten, showed larger effect sizes in acquiring PA than children in 1st grade and above. SES exerted no differential impact on learning PA. However, the language spoken by the children did. English-speaking children showed larger effects of training on PA acquisition than children learning in other languages.

These moderator variables also influenced how much transfer to reading and spelling resulted from PA training. The type of test used to measure reading and spelling influenced effect sizes that were larger on experimenter-devised tests than on standardized tests measuring real word reading and spelling. Effect sizes did not differ on experimenter-devised and standardized pseudoword reading tests.

Properties of training procedures influenced the extent of transfer to reading. Teaching that focused on one or two types of PA manipulations yielded larger effect sizes than teaching three or more PA skills. Teaching children to manipulate phonemes using letters produced bigger effects than teaching without letters. Blending and segmenting instruction showed a much larger effect size on reading than multiple-skill instruction did. Small group instruction produced larger effect sizes on reading than individualized instruction or classroom instruction. Length of training exerted an influence as well, with the lengthiest training associated with the smallest effect size. Classroom teachers provided PA training that was effective in promoting transfer to reading although the effect size of teachers was smaller than the effect size of other trainers. PA training on computers transferred to reading as well.

Characteristics of learners influenced the extent that PA training transferred to reading. Effect sizes on reading were large for at-risk readers while they were moderate for disabled and normally developing readers. Preschoolers exhibited a much larger effect size on reading than did the other grade levels whose effect sizes did not differ. SES made a difference, with mid-to-high SES associated with larger effects than low SES. Also larger effect sizes were evident in reading for English-speaking children than for children speaking other languages.

Analysis of moderator variables as they affected spelling outcomes was complicated by the fact that PA training did not help disabled readers improve in spelling and the pool of spelling effect sizes for disabled readers was homogeneous, indicating that further analyses using moderators was not necessary to explain the result. The effects of moderators were re-analyzed with disabled readers removed from the database. Conclusions regarding the effects of moderator variables on spelling outcomes thus centered on the nondisabled readers.

The only characteristic of PA training that influenced spelling outcomes for nondisabled readers was the use of letters. Children who were taught to manipulate phonemes with letters benefited more in their spelling than children whose manipulations were limited to speech. Whether instruction focused on one or two skills or on multiple skills did not influence spelling in nondisabled readers. Instruction delivered to individuals was as effective as instruction delivered to small groups, and both were more effective than classroom-size groups. The length of training exerted no differential impact on spelling outcomes. Whether the trainer was a teacher or a researcher made no difference. Characteristics of learners did make a difference. Kindergartners benefited more in their spelling than did 1st graders. Students classed as mid-to-high SES showed a larger effect size in spelling than low SES students. PA training in English produced a larger effect on spelling than PA training in other languages.

Features of the design of experiments were related to effect sizes. Findings indicated that rigorous designs yielded strong effects. The majority of the studies used random assignment, and their effect sizes on PA and reading outcomes ranged from moderate to large. About one-third of the studies checked on whether trainers remained faithful to treatment procedures. Effect sizes in these studies were significant and moderate in size. Some studies compared PA treatment groups to control groups that were given some other treatment while other studies used untreated control groups. Neither type of control group consistently produced larger effect sizes. Failure to find larger effects for untreated than for treated control groups indicates that Hawthorne effects did not inflate effect sizes. Studies using smaller samples of children tended to have larger effect sizes than studies using larger samples, a finding consistent with other meta-analyses. However, even in the largest samples, effect sizes were positive and significant.

The Panel also assessed the relationship between methodological rigor and effect size by applying Troia's (1999) criteria to the NRP studies. On PA outcomes, the best designed studies produced the largest effect sizes on all five measures of rigor. On reading outcomes, effect sizes associated with the most rigorous level were close to the largest, if not the largest, effect sizes on four out of five measures: two internal validity measures, one external validity measure, and the overall ranking of rigor. This indicates that the better designed studies produced larger transfer effects in reading than the weaker studies. In sum, findings show that larger effect sizes did not arise mainly from weaker studies that were flawed by threats to internal and external validity.

Interpretations and Issues

Results of the experimental studies allow the Panel to infer that PA training was the cause of improvement in students' phonemic awareness, reading, and spelling performance following training. These findings were replicated multiple times across experiments and thus provide solid support for causal claims. However, results of the analysis of moderator variables rest on more tentative ground. Assessing features of the studies that were associated with stronger or weaker effect sizes is at root a correlational endeavor and thus precludes strong inferences about cause. The primary difficulty is that a third unknown factor may lie in the background explaining the relationships observed. Although findings are suggestive, any conclusions must remain tentative because multiple explanations are possible. In this section, potential misinterpretations of the findings and issues needing further attention from researchers are considered.

The studies in the NRP database included investigations of children at risk for future reading problems as well as children low in SES. However, contrary to the common view that the criteria for identifying at-risk readers includes being economically disadvantaged, authors of the studies investigating at-risk readers 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 these studies, research findings show that one of the two best predictors of reading success is phonemic awareness (Share et al., 1984), so selecting at-risk readers by measuring their PA makes sense. However, because the training targeted this skill, large effect sizes may be less surprising.

The fact that studies in the NRP database departed from the common conception of what it means to be at risk serves to reconcile discrepancies between results for at-risk readers and results for low SES readers. The Panel found that at-risk children showed large effect sizes in acquiring PA ($\underline{d} = 0.95$) and in transferring these skills to reading ($\underline{d} = 0.86$) and spelling ($\underline{d} = 0.76$). Low SES children also showed large effect sizes in phonemic awareness ($\underline{d} = 1.07$) and spelling ($\underline{d} = 0.76$), but only a moderate effect size in reading ($\underline{d} = 0.45$). Smaller effect sizes in reading among low SES children than among at-risk children is explained by the fact that the majority of the at-risk children were not low in SES. Based on these findings, one would expect at-risk children who are both low PA and low SES to exhibit large gains in PA and spelling as a result of PA training but to exhibit moderate gains in reading.

It is noteworthy that low SES children were found to benefit as much from PA training as middle-to-high SES children in acquiring phonemic awareness. This runs counter to Dressman (1999) who argues that low SES children will 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. Dressman bases his expectations on studies showing that low SES children perform more poorly on PA tests than middle-class children. He ignores evidence examining how much low SES children gain in PA when they receive training. According to the NRP findings, low SES children can benefit as much from training as middle-to-high SES children, despite being phonologically or culturally different from the trainers.

One very striking finding was that in contrast to at-risk and normally developing readers, disabled readers' spelling did not benefit at all from PA training. Various reasons for this can be entertained. Other studies have found that disabled readers have special difficulty learning to spell (Bruck, 1993). Perhaps processing difficulties associated with being reading disabled make spelling especially hard to learn. Alternatively, perhaps PA training fails to help older disabled readers with their spelling because the types of words that are spelled in higher grades require knowledge of spelling patterns rather than phonemic segmentation and knowledge of individual letter-sound correspondences. Effects of PA training on spelling may be limited to less complex words that are more phonemically transparent, those taught to beginning readers.

According to NRP findings, children who received training that focused on one or two PA skills 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 focused instruction resulted in more students mastering the skills that were taught. Perhaps teaching multiple skills created some confusion about which manipulations to apply in the reading transfer tasks, or perhaps it obscured children's grasp of the alphabetic principle. Clarifying why multiple skills instruction might limit children's gains in PA and reading needs further study. However, the findings suggest that when multiple

PA skills are the objective, it is prudent to teach one at a time until each is mastered before moving on to the next, and to teach students how each skill applies in reading or spelling tasks.

More important than the number of PA skills to teach is the question of which skills should be taught to children. 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 quite different from the manipulations taught to older students. Easier PA tasks were taught to younger children or to less mature readers while harder PA tasks were taught to older readers. 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. To illustrate, the following tasks are ordered from easy (1) to difficult (6) based on findings of Schatschneider, Francis, Foorman, Fletcher, and 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 the illustrative studies described below, tasks that are appropriate to teach at different grade and reader levels can be seen. The final decision about which PA manipulations to teach should take account of several factors, not only task difficulty, but also whether or not students can already perform the manipulations being taught as determined by pretests, and the use that students are expected to make of the PA skill being taught. The reason to teach first-sound comparisons is to draw preschoolers' or kindergartners' attention to the fact that words have sounds as well as meanings. A reason to teach phoneme segmentation is to help kindergartners or 1st graders generate more complete spellings of words. The reason to teach phoneme blending is to help 1st graders decode words.

One surprising finding in the analysis involved the relationship between training time and outcomes. Effect sizes were larger when PA instruction lasted between 5 and 18 hours than when either less or more time was spent training students. However, caution is needed in interpreting this finding because multiple explanations are possible. Perhaps the goals of instruction were more complex in longer programs. Perhaps the students receiving instruction were harder to teach. Perhaps spending many hours in PA training deprived students of the reading instruction benefiting control groups. Perhaps PA instruction is valuable mainly in helping children achieve basic alphabetic insight. Going beyond this by adding further nuances or complexities may erode learning by producing confusion or boredom. These are only some of the possible reasons why longer training sessions might have produced smaller effect sizes. Questions regarding the optimum length of PA training and factors determining optimum length invite further research. However, two conclusions seem self-evident: that length of training should be regulated by how long it takes students to acquire the PA skills that are taught and that the NRP findings should not be translated into any prescriptions regarding how long teachers should spend teaching PA.

One important moderator variable that was not considered in the analysis is dialect because none of the studies paid attention to this variable. However, regional differences at the phonemic level of language are likely to be important. For example, vowel phoneme categories are not the same across the United States. Some dialects make more phonemic distinctions among vowels than other dialects. Vowels in the three words, <u>marry</u>, <u>Mary</u>, and <u>merry</u> are pronounced identically in some areas of the West but differently in some areas of the East. As a result, no generalizations about these vowel phonemes will suit everyone receiving PA instruction. Another dialectal difference involves preserving or deleting the final consonants in words, for example, past-tense markers such as the /t/ in <u>looked</u>. More research on the impact of dialectal variations on PA learning is needed. The fact that regional phonemic variations exist means that teachers implementing PA training programs need to be aware of their students' dialects and whether they deviate from the phonological systems that are assumed in the programs. Ignoring deviations is likely to undermine the credibility of the instruction.

Another variable related to students' phonological systems but neglected in the analysis is whether English is the first or second language of students. The problem here is that phonemes in English may not be phonemes in ESL students' first language. To understand this requires distinguishing between phonemes and phones. Phonemes are the smallest units in speech that signal a difference in meaning to a listener who knows the language. Phones are also the smallest units in speech but are described by acoustic and articulatory properties. To perceive phonemes, speakers use categories that were constructed in their minds when they learned their particular language. In contrast, phones are defined by their physical properties. Phonemes are broader categories that may include several phones, called allophones, differing in their articulatory features. Even though the allophones differ, speaker/listeners process them as the same phoneme. For example, the initial sounds in chop and shop are articulated differently, so they are two different phones. 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 different phones are the same phoneme. 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 that Chinese and Japanese speakers process /l/ and /r/ as the same phoneme in English words.

The distinction between phonemes and phones may seem trivial, but it is not. If teachers have students who are learning English as a second language, they need to realize that their students are almost bound to misperceive some English phonemes because their linguistic minds are programmed to categorize phonemes in their first language, and this system may conflict with the phoneme categorization system in English. Their confusions will be most apparent when they select letters to spell unfamiliar words. If they know Spanish, they may select CH when they should use SH. If they know Japanese or Chinese, they may confuse L and R. When teachers teach PA, they need to be sensitive to these sources of difficulty faced by their ESL students.

The Role of PA in Reading Acquisition Processes

Findings of the meta-analyses show that PA training benefits the processes involved in reading real words, pseudowords, and text reading. It also benefits spelling skills in normally progressing readers below 2nd grade and in beginners at risk for developing reading problems. There are several reasons why PA training is thought to help children learn to read and spell.

The English writing system is alphabetic. Breaking the code entails figuring out how graphemes represent phonemes. These relationships, though systematic, are variable across word spellings. The same letters may symbolize more than one phoneme, and single phonemes may be represented by alternative graphemes. The vowels are especially variable. This lack of transparency makes it harder for beginners to figure out the system without help.

Speech is seamless and has no breaks signaling where one phoneme ends and the next begins. Also, phonemes overlap and are coarticulated, which further obscures their separate identities. Another barrier to developing PA is that speakers focus their attention on the meanings of utterances, not on sounds. Unless they are trying to learn an alphabetic code, there is no reason to notice and ponder the phonemic level of language. These facts explain why beginners have difficulty acquiring PA and why they benefit from explicit instruction in PA.

An essential part of the reading process involves learning to read words in various ways (Ehri, 1991, 1994). Because phonemes in words correspond to graphemes in the English writing system, all of these ways of reading words are easier to acquire when beginners possess PA. Phoneme identity is needed to attach phonemes to letters for reading and spelling words. The skill of blending is needed to decode unfamiliar words. Being able to segment and blend onsets and rimes in words helps children read unfamiliar words by analogy to known words. Phonemic segmentation helps children remember how to read and spell words because it helps them distinguish the phonemes that are bonded to graphemes when a word's written form is retained in memory. When unfamiliar words are read in text, students may apply decoding skills, or they may combine grapheme-phoneme cues with meaning cues to derive the word (Tunmer & Chapman, 1998).

It is important to note that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping children understand and use the alphabetic system to read and write. This is why including letters in the process of teaching children to manipulate phonemes is important. PA training with letters helps learners determine how phonemes match up to graphemes within words and thus facilitates transfer to reading and spelling.

It is important to recognize that children will acquire some 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. This is indicated by evidence that people who do not learn to read in an alphabetic system do not develop PA (Mann, 1987; Morais et al., 1987; Read et al., 1987). It is also indicated by the fact that, in many of the studies reviewed, control groups showed improvement in phonemic awareness from pretests to posttests, very likely because of the reading and writing instruction they received in their regular classrooms. However, the extent of PA needed to contribute maximally to children's reading development does not arise from incidental learning or instruction that is not focused on this objective. This is indicated by the finding that children receiving explicit training in PA gained much more PA and reading skill than children in the control groups.

It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness; therefore, PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling while others may know only a few letters and have no reading skill. The nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA prior to beginning PA instruction. This will indicate which children need the instruction and which do not; which children need to be taught rudimentary levels of PA, for example, segmenting initial sounds in words; and which need more advanced levels involving segmenting or blending with letters.

In the rush to teach phonemic awareness, it is important not to overlook the need to teach letters as well. The NRP 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 lower-case letter shapes, names, and sounds to learn. The shapes of many letters are similar, and, therefore, easily confused with one another. Letter learning requires retaining shapes, names, and sounds in memory and, in fact, overlearning them so that letters can be processed automatically in reading and writing words (Adams, 1990). Thus, to ensure that instruction in phonemic awareness is effective, it needs to include instruction in graphemes as well as 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 make the connection between the PA skills taught and their application to reading and writing tasks. In most of the studies reviewed, researchers did not do this when they presented the transfer tasks to students following training. Despite this, significant and sizable transfer effects were observed. In a study by Cunningham (1990), who did examine application effects, students in one group not only were taught to segment and blend but also were shown how to apply these skills in reading words. Another group received the same PA training but not the application training. Effect sizes on reading outcomes were much larger when 1st graders received the application instruction than when they did not. This suggests that results of the NRP meta-analysis actually underestimate the magnitude of effects that would result if children received explicit instruction and practice in applying PA skills in their reading and writing.

It is important to note that when PA is taught with letters, it qualifies as phonics instruction. When PA training involves teaching students to pronounce the sounds associated with letters and to blend the sounds to form words, it qualifies as synthetic phonics. When PA training involves teaching students to segment words into phonemes and to select letters for those phonemes, it is the equivalent of teaching students to spell words phonemically, which is another form of phonics instruction. These methods of teaching phonics existed long before they became identified as forms of phonemic awareness training (Balmuth, 1982; Chall, 1967). Although teaching children to manipulate sounds in spoken words may be new, phonemic awareness training that involves segmenting and blending with letters is not. Only the label is new. Explicit instruction in the alphabetic principle necessarily includes attention to phonemes because these are the phonological units that match up to letters. According to NRP findings, it is likely that the inclusion of phonemic awareness training in phonics instruction is a key component contributing to its effectiveness in teaching children to read.

It is important to note that various approaches to beginning reading instruction may provide at least some phonemic awareness training although it may not be presented systematically or thoroughly enough to maximize its contribution to reading and writing. Whole language instruction that teaches students to invent spellings by detecting phonemes in words and representing them with letters offers a form of PA training. In Reading Recovery© (RR), students may acquire phonemic awareness through the spelling instruction they receive (Clay, 1985). Three studies in the database compared outcomes of standard whole language instruction, or RR instruction, to outcomes of the same instruction with PA training added (Castle et al., 1994; Hatcher et al., 1994; Iversen & Tunmer, 1993). Overall effect sizes were variable ranging from negative to large positive (see Appendix and illustrative studies below). One factor possibly limiting outcome differences between treatment and control groups is the extent to which control students acquired PA from the instruction they received. Although whole language programs and RR programs include some phonemic awareness training, findings of the NRP meta-analysis indicate that strengthening the training offered in spelling activities by making it more systematic, thorough, and explicit, is likely to improve these programs' success in helping children learn to read and spell.

Classroom Instruction in PA: Some Illustrations

NRP findings show that PA training programs implemented by teachers in classrooms are effective in teaching phonemic awareness to students, and this training boosts children's reading and spelling performance. To identify characteristics of programs that were used successfully by classroom teachers, the Panel examined a few illustrative studies selected from a total of 15 (Blachman, Ball, Black, & Tangel, 1994; Brady, Fowler, Stone, & Winbury, 1994; Brennan & Ireson, 1997; Bus, 1986; Haddock, 1976; Kennedy & Backman, 1993; Kozminsky & Kozminsky, 1995; Lie, 1991; Lundberg, Frost, & Petersen, 1988; McGuinness, McGuinness, & Donohue, 1995; O'Connor, Notari-Syverson, & Vadasy, 1996; Olofsson & Lundberg, 1983; Schneider, Kuspert, Roth, Vise, & Marx, 1997; Tangel & Blachman, 1992; Williams, 1980).

One 8-month-long, carefully structured program for kindergartners was developed and tested by Lundberg, Frost, and Peterson (1988). 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 followed by 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 who spoke peculiarly, syllable by syllable, and they figured out what he said. Phoneme analysis was introduced in the 3rd month by having children identify phonemes in initial positions of words, mainly continuants and vowel sounds which are easy to stretch out and hold. The teacher helped children find the sounds by stretching them, for example, "Mmmmmark" or by repeating the stop consonants that cannot be held, for example, "T-T-T-Tom." Children also practiced adding and deleting phonemes from words. In the 5th month of the program, phoneme segmentation and blending were introduced, first with two-phoneme words and then longer words. Many of the activities were designed for children's enjoyment and consisted of dancing, singing, and other noncompetitive social games.

Teachers were trained in an inservice course that provided theoretical background as well as videotaped examples of training sessions. They practiced and refined the skills necessary to teach the program during the year prior to implementing it. Teachers of the control group followed the regular preschool program, which emphasized social and aesthetic aspects of development rather than cognitive and linguistic aspects. Treatment and control schools were located in geographically distant parts of Denmark.

The Danish program was adapted and tested by other researchers including Schneider et al. (1997) who taught PA to German kindergartners. His study included two experiments and a total of 22 teachers who taught PA in the treatment conditions. Control groups received the regular kindergarten curriculum. The second experiment was conducted to improve on the first. Teacher training was less extensive in the German study than in the Danish study. It lasted 2 months and included theoretical background and tutoring sessions in which teachers practiced the games and exercises and received feedback.

In both the Danish and German studies, training produced large effect sizes on the acquisition of phonemic awareness, ranging from 0.70 to 0.82. Effect sizes on reading outcomes were small to moderate when measured the following year in 1st grade: d = 0.19 (Denmark), d = 0.26 and 0.45 (Germany).

An adaptation of the Danish program was tested with English-speaking kindergartners by Brennan and Ireson (1997). However, only one teacher and her class of 12 students formed the PA treatment group, which was compared to one no-treatment control class. Although this is a weaker design yielding less reliable findings, the effect size was impressive. The impact of training on word reading was large, with an effect size of d = 1.17. This provides some evidence that the Danish program can be used effectively in American classrooms. A translation of the program has been published (Adams, Foorman, Lundberg, & Beeler, 1998). Whether teachers need further help beyond the manual to implement the program effectively with their students needs to be studied.

The Danish program did not include letter manipulation. However, the meta-analysis showed that when PA is taught with letters, it is more effective. A program for kindergartners that included letters was developed and tested by Blachman and her colleagues (Ball & Blachman, 1991; Blachman et al., 1994; Tangel & Blachman, 1992). Blachman et al. (1994) taught 10 teachers and their teaching assistants to deliver PA training to low-income, inner-city kindergartners. Children were taught in groups of four or five for 15 to 20 minutes per day, 4 times each week. The program lasted 11 weeks. The teachers were trained in seven 2-hour inservice workshops, during which they were taught a theoretical framework; they practiced instructional activities; and they asked questions about ways of implementing 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 to represent phonemes in three-phoneme words. A variety of games was used to reinforce grapheme-phoneme correspondences. The control group in this study followed a traditional kindergarten curriculum that included instruction in letter names and sounds. Results of the study were very positive. Children receiving PA training outperformed controls on PA tasks, with an effect size of d = 1.83, and training transferred to reading, d = 0.65, and to spelling, d = 0.94.

Another program in the NRP set of studies was administered by teachers to small groups of older disabled readers. Williams (1980) developed and tested the ABDs program, which taught students ages 7 to 12 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 and then three phonemes. Then blending was applied to the same words. Children performed manipulations with wooden markers at first 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 to the set later in the program. Words with consonant clusters were introduced. Finally two-syllable words were added. 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 presented and 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 evaluated the ABDs program again the following year, this time not with volunteer teachers but with 20 teachers who were mandated to use the program. They completed on average 6.6 units, about half the program. The 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 training. Effect sizes were large, $\underline{d} = 1.05$ for the 1st year, and $\underline{d} = 0.97$ for the 2nd year. This indicates that the ABDs program was highly effective at teaching decoding skill to disabled readers.

Other Programs to Teach PA

Various programs were used to teach PA across studies. Presenting descriptions of these programs serves to clarify how studies in the database were structured and the variety of ways that PA was taught. Some programs had special features that enhanced their effectiveness. In the study by Cunningham (1990), one treatment group was taught metacognitive skills along with PA. Cunningham worked with normally progressing readers in kindergarten and 1st grade. A puppet was utilized to interact with children. PA training was limited to the oral mode, with no letter-sound instruction. Training was conducted in small groups for 10 weeks. Three treatments were compared. One treatment group received PA training in segmenting and blending phonemes. Another group received a somewhat abbreviated version of this training and spent the extra time in metacognitive activities that included learning about the goals and purposes of each PA manipulation, reviewing how that lesson related to previous lessons, and observing and practicing how to use the skill for reading. The control group spent equal time engaged in a story listening treament.

Results showed that at the end of PA training, the two treatment groups outperformed the control group on measures of PA and reading in both grades. In addition, 1st graders who had received both PA and metacognitive training achieved higher reading scores than 1st graders receiving only PA training. One possible reason why the advantage was limited to 1st grade is that 1st graders, but not kindergartners, were receiving formal reading instruction concurrently in their classrooms, so they had a chance to apply their PA knowledge on a daily basis. In fact, some 1st graders told the experimenter that they used what they had been taught to decode words in their classroom reading groups. These findings indicate that a metacognitive component may be valuable in providing a bridge between PA skills and reading processes. This may be particularly true in PA programs that do not teach phoneme manipulation with letters.

The ADD program (Auditory Discrimination in Depth) was developed by Lindamood and Lindamood (1975) to teach PA. The unique feature of this program is that it teaches children to identify and monitor articulatory gestures associated with phonemes. As already discussed, phoneme segmentation is difficult because there are no boundaries in speech telling us where one phoneme ends and the next begins. Rather phonemes are coarticulated to produce speech without any seams. One very helpful 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 <u>meat</u> and these are reflected in three successive mouth movements: your lips closing for /m/, your lips opening into a smile shape for the vowel, then your tongue tapping the roof of your mouth for /t/. Pictures of mouth positions can be used to help children distinguish phonemes in pronunciations of words. Also, mirrors help children explore what their own mouths are doing when they pronounce words.

Four studies in the NRP database implemented the ADD program to teach PA (Kennedy & Backman, 1993; McGuinness et al., 1995; (Wise, King, & Olson, 1999; Wise, King, & Olson, in press). Children received extensive training discovering and categorizing the various phonemes in English by analyzing their own mouth movements, often using mirrors. They learned to label these sounds, for example, lip poppers, tip tappers, and scrapers. They learned to track 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 1.22 for 1st graders (McGuinness et al., 1995) to 0.15 for older disabled readers (Wise, King, & Olson, 1999).

An example of a program focused on teaching only one type of phoneme manipulation was that studied by Byrne and Fielding-Barnsley (1991) for preschoolers, called Sound Foundations. This program taught phoneme identity. Children learned to recognize instances of the same sound in both initial and final positions across different words. The following sounds received primary attention: /s/, /š/ as in ship, /l/, /m/, /p/, /t/, /g/, /ae/ as in bat, /ɛ/ as in bet. Children were shown several large posters covered with pictures of objects. Their job was to pick out from a larger set the objects having a specified beginning or ending sound, for example, sea, seal, sailor, sand. 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, preschoolers averaging 4.5 years of age received either the PA training described above or control training that focused on story reading and semantic activities with the same posters and worksheets. Children were trained in groups of 4 to 6 children, one 30-minute lesson per week for 12 weeks. At the end of training, children in the PA-trained group were able to identify substantially more initial and final phonemes in words than control students. They demonstrated superior skill identifying not only sounds they had practiced but also unpracticed sounds, indicating that phoneme identity skill transferred to untaught phonemes. These researchers also gave students a simplified word reading task in which children were shown a word and identified it from two spoken choices (e.g., "Does this [sat] say sat or mat?"). Trained students read more words than control students, indicating that PA training improved preschoolers' rudimentary word recognition skill.

These researchers also investigated the long-term impact of PA training (Byrne & Fielding-Barnsley, 1993, 1995). Children were tested during the next 3 years in school. At the end of kindergarten, trained children were only slightly superior to controls in PA, indicating that learning to read had narrowed the gap in PA between the two groups. At the end of each successive grade, the PA-trained group read significantly more pseudowords than controls, indicating that PA training benefited children's decoding skill. At the end of 2nd grade, there was a marginal difference in reading comprehension favoring the PA-trained students. However, the 2nd graders did not differ in reading real words or in spelling words.

One possible reason why long-term training effects were not stronger in this study is that the formal reading and spelling instruction that children received in school was sufficiently effective to compensate for the advantage provided by preschool training in PA. Also, the PA training that students received was focused rather than comprehensive and amounted to only 6 hours total. It may take a more comprehensive and extensive training program to exert stronger long-term effects.

The effectiveness of different ways to teach PA was examined by O'Connor et al. (1995), who inquired whether PA training has to be broad rather than focused to be most effective. They selected at-risk kindergartners with low PA and randomly assigned them to one of three training conditions. In the comprehensive treatment, children performed a variety of sound manipulation activities that included isolating, segmenting, blending, and deleting phonemes; segmenting and blending syllables and onset-rime units; and working with rhyming words. In the focused treatment, children practiced segmenting and blending onsets, rimes, and phonemes only. Training extended for 10 weeks, two 15-minute sessions per week, totaling 5 hours. Beginning in the 5th week, letter-sound associations were taught for the sounds being practiced orally in both groups. However, children were not taught how to use letters to manipulate phonemes in the PA activities. The third treatment, a control condition, received only the letter-sound instruction.

Comparison of phonemic awareness following training showed that the treated groups performed equally well and both outperformed controls, indicating that both types of training were equally effective in teaching PA. To measure transfer to reading, a simplified word learning task was devised. After children learned to associate four letters and sounds, they were given practice learning to read five words composed of the letters and sounds: am, at, mat, sat, sam. Each word was taught by saying, "This is aaaaat, at." Results revealed that only the focused group learned to read the words in fewer trials than the control group, not the comprehensive group. This suggests that concentrating instructional time on segmenting and blending may contribute more to reading skill than diverting attention to many PA activities. These findings are consistent with those in the NRP meta-analysis indicating the greater impact of segmenting and blending than multiskill instruction on reading outcomes. One might question the use of a simplified word reading task to draw inferences about general reading acquisition. However, these kindergartners were beginning readers so that more advanced reading tests would have been too difficult.

The separate and combined contributions of instruction in segmentation and blending were examined by Davidson and Jenkins (1994), who gave kindergartners with low PA one or another of four types of training. In the segmentation treatment, each word was pronounced, and children were taught to say its separate sounds. In the blending treatment, children listened to the separate sounds and learned to blend them into words. In the segmentation-and-blending treatment, children learned first to segment, then to blend the words. In the control condition, children listened to stories. Children were taught to a criterion of mastery. The words and nonwords analyzed during training had two phonemes formed out of continuant consonants and long vowels (e.g., my, vo, low, way). At the end of training, all students were taught eight letters for the sounds that treatment groups had practiced. Then two literacy tests were given in which children practiced and received feedback in learning to read and learning to spell two-phoneme words. These words were formed from the same letter-sounds but they had not been taught during training.

Results showed that the groups learned the PA skill that they were taught but performed poorly on the untaught skill. This indicates that teaching students either segmentation or blending does not improve their performance in the other skill. On the measures of reading and spelling, both the segmentation and combination groups performed similarly and outperformed the control group. However, the blending group did not do better than the control group. This indicates that teaching beginners to segment is as effective for learning to read words as teaching beginners to segment and blend. In contrast, teaching beginners only to blend is not effective. These findings were replicated in a similar study by Torgesen et al. (1992).

Although blending made a poor showing in these studies, Reitsma and Wesseling (1998) reported more success in a study with kindergartners in the Netherlands. They used a computer to teach kindergartners how to blend three-phoneme Dutch words (e.g., lief, geit, met). No limits were placed on the variety of phonemes in the words. All phoneme manipulations were conducted in speech without any letters. First, children were taught a set of vocabulary words, and then these were used in various blending exercises. Children listened to a sequence of segmented sounds, and then clicked on the picture corresponding to that word. Children listened to two successively segmented words and clicked "same" or "different." Children listened to words, either pronounced as wholes or segmented, and then had to find which of several boxes on the screen contained the other form of the word. If a whole word was heard, they had to find its segmented form. If a segmented word was heard, they had to find its whole form. In all these exercises, the incorrect word choices differed by several phonemes from the correct choice for some items but only by one phoneme for other items, making processing more difficult. In the control group, children completed vocabulary exercises on the computer.

At the end of kindergarten, PA tests of children's ability to blend and to segment words revealed superior blending performance by the trained group over the control group, but no difference in segmentation performance. Thus, training effects were limited to blending which was the skill taught, and blending skill did not transfer to segmentation. The following year, in 1st grade, children's ability to read words was examined. Long-term effects of the blending exercises were evident because trained children read more words than control children. However, no effects on spelling were detected. These results suggest that extensive training to develop blending skills does benefit reading acquisition. Blending is thought to contribute to reading by enabling children to decode new words they have not yet learned to read. Also, findings indicate the effectiveness of using computers to teach PA to kindergartners.

One instructional activity that is maximally effective for teaching PA in a way that builds a bridge to reading and spelling is that of teaching children to invent phonemically more complete spellings of words. Typically, kindergartners who know letter names or sounds can represent the more salient sounds in words such as beginning and ending sounds, for example, writing B to spell beaver or R to spell arm. Sometimes their spellings are not conventional, for example, writing Y to spell wife. However, the important achievement is that they can distinguish sounds in words. Once they can do this, then teachers can help them detect additional sounds in words and learn conventional spellings for those sounds.

In a study by Ehri and Wilce (1987b), kindergartners were taught individually how to generate phonemic spellings of words and nonwords by segmenting words into phonemes and selecting letters representing those phonemes. Children who qualified for the study could already name the six consonant and four vowel letters that were used in training. All names contained the relevant sounds in their names (T, S, N, L, K, P, A, E, I, O).

Instruction began with two-phoneme words and nonwords and progressed to three-phoneme words and words with consonant clusters. Children were helped to break words into phonemes by directing their attention to articulatory gestures. They were helped to select letters by focusing on sounds in letter names. They mastered shorter words before advancing to longer words. Children in the control group practiced matching the ten letters and sounds in isolation. Articulatory gestures and letter names were used to correct their errors as well. On posttests after training, effect sizes were large on measures of segmentation and spelling. The measure of reading involved giving children practice learning to read 12 similarly spelled words for several trials. The words were spelled phonemically with the letter-sounds taught, for example, SEL (seal), SNAK (snake), SLIS (slice). The effect size was large, $d = 0.97$. These findings indicate that teaching children to segment and spell helps them learn to read as well as spell words.

In many PA training studies, the instructional context was not considered. However, there were some exceptions. Iversen and Tunmer (1993) incorporated PA training into Clay's (1985) Reading Recovery© program to examine whether systematic instruction in PA would make the program more effective. At-risk readers in 1st grade were assigned to one of three groups, a group receiving standard Reading Recovery© instruction, a group receiving modified RR instruction, and an alternative, non-RR intervention group. In the modified RR treatment, after children had learned most letters, they manipulated magnetic letter forms to make, break, and build new words having similar spellings and pronunciations, for example, reading and and then changing it to hand, sand, band. Training progressed from initial sounds to final sounds and then to medial sounds. Children added, deleted, and substituted letters in their manipulations and also read the changed words. Later, the task becomes a writing rather than a manipulation task.

Findings showed that both forms of RR enabled children to reach prescribed reading levels that qualified them to exit the remedial program. However, children who received modified RR attained prescribed levels more quickly than children receiving the standard program (i.e., a mean of 41.75 lessons for modified RR vs. 57.31 lessons for standard RR). This indicates that adding PA training improved RR by increasing its efficiency. At the end of training, however, both groups performed at very similar levels on PA outcomes and reading outcomes, indicating that both forms of the RR program enabled children to attain similar levels of PA and reading. On followup tests given at the end of the school year, performance of the groups remained very similar.

Hatcher et al. (1994) also examined whether adding PA training to a Reading Recovery© program would improve its success. The participants were 7-year-old poor readers. The PA training that was added to RR involved teaching children to perform different types of PA, including segmentation, blending, deletion, substitution, and transposition of phonemes. Children also practiced linking letters to phonemes in various spelling and writing tasks. Effect sizes, though small, favored the PA-trained group ($\underline{d}$ = 0.24 for PA, $\underline{d}$ = 0.31 for reading and spelling).

Castle et al. (1994) examined the contribution of PA training to reading acquisition in a whole language program. Kindergartners with low PA were assigned to treatment and control groups. PA training included segmentation, blending, substitution, and deletion. Letters were incorporated into the PA activities later in the program. Two control groups were included, one receiving an alternative, unrelated treatment and another receiving no treatment other than the whole language instruction provided to all participants in their classrooms. Results showed that the PA-trained group spelled more words and decoded many more pseudowords than the two control groups. However, the groups did not differ in reading real words or in reading connected text. These findings indicate that adding PA instruction to a whole language program enhances students' decoding and spelling skills but not their other reading skills.

Wise et al. (in press) evaluated the effects of PA training against training that taught children reading comprehension strategies and gave them extensive text reading practice on computers. The children were 200 disabled readers in grades 2 to 5. Both treatment and control groups spent time reading stories on the computer. They could touch any unknown word with a cursor and have it identified. Comprehension questions were answered periodically. Controls spent extra time reading on the computer while the PA-trained group completed various types of PA activities administered by the computer. For example, the computer asked the child to show <u>feef</u>. The child selected and ordered letter-sound symbols with a mouse. Synthetic speech pronounced whatever the child assembled, and the child continued to manipulate letters until achieving a match. Then the computer asked the child to change the word to <u>feem</u>. Lessons began with two-phoneme words and progressed to longer words. There were several other PA activities besides this one.

On the posttests, PA-trained children outperformed controls on tests of phonemic awareness and pseudoword reading tests. Also, they read more words when there were no time constraints. However, controls displayed superior time-limited word reading. Both groups made similar gains in spelling and reading comprehension. Interestingly, when the analysis of word reading took account of grade level, 2nd graders gained more than older children and they showed a much greater advantage for PA training over the control training than did older children. These findings suggest that PA training may be more beneficial to younger than to older disabled readers.

In sum, these illustrative studies enrich the understanding of the data contributing to the NRP meta-analysis. They show that various types of instruction were utilized to teach PA at various grade levels. They show how different studies were designed and the nature of their findings. Also, they draw attention to other potentially important features that were not addressed in the meta-analysis because of an inadequate number of cases.

Implications for Reading Instruction

1. Can phonemic awareness be taught, and does it help children learn to read and spell?

Results of the meta-analysis showed that teaching phonemic awareness to children is clearly effective. It improves their ability to manipulate phonemes in speech. This skill transfers and helps them learn to read and spell. PA training benefits not only word reading but also reading comprehension. PA training contributes to children's ability to read and spell for months, if not years, after the training has ended. Effects of PA training are enhanced when children are taught how to apply PA skills to reading and writing tasks.

2. Which students benefit in their reading?

Teaching phonemic awareness helps many different students learn to read, including preschoolers, kindergartners, and 1st graders who are just starting to learn to read. This includes beginners who are low in PA and are thus at risk for developing reading problems in the future. This includes older disabled readers who have already developed reading problems. This includes children from various SES levels. This includes students who are taught to read in English, as well as students taught to read in other alphabetic languages.

3. Which students benefit in their spelling?

Teaching phonemic awareness helps preschoolers, kindergartners, and 1st graders learn to spell. It helps children at risk for future reading problems also. It helps low as well as middle-to-high SES children. It helps students learning to spell in English as well as students learning in other languages. However, PA training is ineffective for improving spelling in reading-disabled students. This is consistent with other research indicating that disabled readers have a hard time learning to spell.

4. Which methods of teaching PA work best in helping children acquire phonemic awareness?

Various forms of phoneme manipulation might be taught, including identifying or categorizing the phonemes in words, segmenting words into phonemes, blending phonemes to form words, deleting phonemes from words, or manipulating onsets and rimes in words. In some programs, only one PA skill is taught, while in other programs, two or more skills are combined. Some programs teach children to use letters to manipulate phonemes and others limit training to speech. All of these approaches appear to be effective for helping children learn to manipulate phonemes. Focusing on one or two skills produces larger effects than a multiskilled approach. Teaching PA with letters helps students acquire PA more effectively than teaching PA without letters.

5. Which methods of teaching PA have the greatest impact on learning to read?

Although all of the approaches exert a significant effect on reading, instruction that focuses on one or two skills produces greater transfer than a multiskilled approach. Teaching students to segment and blend benefits reading more than a multiskilled approach. Teaching students to manipulate phonemes with letters yields larger effects than teaching students without letters, not surprisingly because letters help children make the connection between PA and its application to reading. Teaching children to blend the phonemes represented by letters is the equivalent of decoding instruction. Being explicit about the connection between PA skills and reading also strengthens training effects.

6. Which methods of teaching PA have the greatest impact on learning to spell?

Teaching PA helps nondisabled readers below 2nd grade learn to spell. Methods that teach children to manipulate phonemes with letters are more effective than methods limiting manipulation to spoken units. Teaching children to segment phonemes in words and represent them with letters is the equivalent of invented spelling instruction.

7. How important is it to teach letters as well as phonemic awareness?

It is essential to teach letters as well as phonemic awareness to beginners. PA training is more effective when children are taught to use letters to manipulate phonemes. This is because knowledge of letters is essential for transfer to reading and spelling. Learning all the letters of the alphabet is not easy, particularly for children who come to school knowing few of them. Shapes, names, and sounds need to be overlearned so that children can work with them automatically to read and spell words. Thus, if children do not know letters, this needs to be taught along with PA.

8. How much time is required for PA instruction to be effective?

In the NRP analysis, studies that spent between 5 and 18 hours teaching PA yielded very large effects on the acquisition of phonemic awareness. Studies that spent longer or less time than this also yielded significant effect sizes, but effects were moderate and only half as large. Transfer to reading was greatest for studies lasting less than 20 hours. In fact, effect sizes were more than twice as large for shorter programs than for the longest-lasting programs.

Caution is needed in drawing conclusions from this finding. Although it suggests that less instructional time is better, it ignores reasons why training that lasted longer might have been less effective. Perhaps the PA skills being taught were more complex, or perhaps the learners were harder to teach, or perhaps, as a result of time spent in training, PA-trained students received less instruction in reading than students in the control groups.

The Panel concludes that it is wrong to make any declarations about how long effective instruction in PA needs to last based on the NRP findings. Rather, decisions should be influenced by reason, moderation, and situational factors. The answer depends on the goals of instruction, how many different PA skills are to be taught, whether letters are included, how much or how little the learners already know about PA when they begin, whether they are disabled readers, whether provision is made for facilitating transfer to reading and spelling, and so forth. Individual children will differ in the amount of training time they need to acquire PA. What is probably most important is to tailor training time to student learning by assessing who has and who has not acquired the skills being taught as training proceeds. Children who are still having trouble should continue PA training while those who have learned the skills should move on to other reading and writing instruction.

Not only the total training time but also the length of single training sessions must be considered. In the NRP database, the average length of sessions was 25 minutes. Few sessions lasted more than 30 minutes, and these tended to occur with older disabled readers, not with younger children. From this, the Panel concludes that sessions should probably not exceed 30 minutes in length.

9. Can classroom teachers teach PA effectively to their students?

Classroom teachers are definitely able to teach PA effectively. In the NRP analysis, their effect size on the acquisition of PA was large. The training they provided transferred and improved students' reading and spelling, and the effect on reading continued beyond training. It was not possible to specify the amount of training required to enable trainers to be effective. This relationship was not examined in the studies. Only 15 studies reported the length of training provided to trainers. It ranged from 2 to 90 hours, with a mean of 21 hours. This suggests that the amount of training required may be quite modest and reasonable for inservice instruction.

10. Is instruction most effectively delivered to individual students, to small groups, or to full classrooms of students?

Although individual tutoring is commonly regarded as the most effective unit of instruction, NRP findings indicate that small groups are the best way to teach phonemic awareness to children. Also, small groups facilitate greater transfer to reading than the other two teaching units. This may hold true for several reasons. Children may benefit from observing their peers respond and receive feedback or from listening to their peers' comments and explanations. Or children may be more attentive and motivated to learn so that they do well in the eyes of their peers.

11. Is evidence for the effectiveness of PA training on reading outcomes derived from strongly designed or weakly designed studies?

The NRP analyses show that the evidence rests solidly on well-designed studies. Significant effect sizes were apparent on standardized tests as well as experimenter-designed tests. Random assignment of children to groups yielded significant effects. In fact, this effect size was larger than that for the nonequivalent group design. Studies in which treatment fidelity was checked yielded a moderate effect size. Significant effects occurred not only when PA-trained groups were compared to untreated control groups but also when they were compared to treated controls. Significant effects were detected with larger as well as smaller samples of children. When Troia's (1999) criteria for methodological rigor were applied to studies, the most rigorous studies yielded the largest effect sizes. The Panel concludes that evidence for the effectiveness of PA training on reading outcomes comes from well-designed experiments. In fact, researchers are advised that they have the best chance of observing strong effects if they apply the most rigor in designing their PA studies.

12. Are the results ready for implementation in the classroom?

This section of the NRP report includes many ideas that provide guidance to teachers in designing PA instruction and in evaluating and selecting programs with the best chance for success. However, in implementing PA instruction in the classroom, teachers should bear in mind several serious cautions:

- PA training does not constitute a complete reading program. Although the present meta-analysis confirms that PA is a key component that contributes significantly to the effectiveness of beginning reading and spelling instruction, there is obviously much more that children need to be taught to acquire reading and writing competence. PA instruction is intended only as a foundational piece. It helps children grasp how the alphabetic system works. It helps children read and spell words in various ways. However, literacy acquisition is a complex process for which there is no single key to success. Teaching phonemic awareness does not ensure that children will learn to read and write. Many competencies must be acquired for this to happen.
- Exactly how PA instruction should be taught by teachers in their classrooms is not clearly specified by the research. A variety of programs was found to be effective. The studies are useful in identifying features that are important and should be considered in selecting programs and planning classroom instruction. Ultimately, though, teachers need to evaluate the methods they use against measured success in their own students.
- One factor that is very important to effective classroom instruction but has not been addressed in the PA training research is the extent to which these programs motivate both students and teachers. It seems self-evident that instructional techniques for developing PA need to be relevant, engaging, interesting, and motivating in order to promote optimal learning in children. However, the research has not focused on this factor. Neither has the research examined which techniques are most engaging for teachers. It seems self-evident that teachers are most effective when they are enthusiastic and enjoy what they are teaching. In selecting ways to teach PA, teachers need to take account of motivational aspects of programs for themselves as well as their students.
- Teachers should recognize that acquiring phonemic awareness is a means rather than an end. PA is not acquired for its own sake but rather for its value in helping learners understand and use the alphabetic system to read and write. This is why it is important to include letters when teaching children to manipulate phonemes and why it is important to be explicit about how children are to use the PA skills in reading and writing tasks.
- It is important to recognize that children will acquire some 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. However, incidental instruction that does not focus on teaching PA falls short in its contribution to children's reading and spelling development.
- It is important to recognize that children will differ in their phonemic awareness and that some will need more instruction than others. In kindergarten, most children will be nonreaders and will have little phonemic awareness; so, PA instruction should benefit everyone. In 1st grade, some children will be reading and spelling already while others may know only a few letters and have no reading skill. The nonreaders will need much more PA and letter instruction than those already reading. Among readers in 1st and 2nd grades, there may be variation in how well children can perform more advanced forms of PA, that is, manipulations involving segmenting and blending with letters. The best approach is for teachers to assess students' PA prior to beginning PA instruction. This will indicate which children need the instruction and which do not; which children need to be taught rudimentary levels of PA, for example, segmenting initial sounds in words; and which need more advanced levels involving segmenting or blending with letters.

Directions for Further Research

A large number of experiments have been conducted to test whether phonemic awareness training helps children learn to read. Results have been sufficiently positive to sustain confidence that this treatment is indeed effective across a variety of child and training conditions. However, there are still some questions needing further attention from researchers.

1. Training Teachers to Teach PA

Findings of a few studies have raised doubt that teachers possess sufficient phonemic awareness to teach this skill adequately on their own (Moats, 1994; Scarborough, Ehri, Olson, & Fowler, 1998). These studies indicate that teachers fall short in manipulating phonemes correctly. However, the studies do not show that this lack of knowledge limits teachers' ability to *learn* to teach PA adequately. Results of the Panel's analysis indicate that with training, teachers can teach PA effectively.

Research is needed to clarify what sort of knowledge and training maximizes teachers' effectiveness in teaching PA and in integrating this instruction with beginning reading instruction. This includes both preservice training and inservice training that covers instruction for preschoolers, primary students, and older disabled readers. Questions to be addressed are: How much and what sort of linguistic knowledge about phonemes, graphemes, and the alphabetic system need to be taught to teachers? How much knowledge about literacy learning processes and their course of development in beginning readers needs to be understood by teachers? Teachers may need to know how phonemic awareness develops in children, which tasks are easier and which are harder, what techniques help children focus on phoneme-size units such as monitoring articulatory cues, what kinds of mistakes children commonly make, what the origin is of these mistakes, 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. Such knowledge should help teachers utilize this approach to teach PA. Research is needed to address these possibilities.

2. Use of Small Groups, Large Groups, or Individual Tutoring to Teach PA

In the meta-analysis of instructional programs, size of training unit was uncovered as a property that affected outcomes differentially. Small group instruction was associated with much larger effect sizes than individual or classroom instruction. However, these findings are correlational. That is, differences emerged across studies. Differences did not arise in studies that manipulated this variable experimentally. As a result, attributing cause to this property is highly tentative and open to other interpretations. The next step for researchers is to determine experimentally whether small group instruction is indeed a better way to teach PA than individual and classroom instruction and, if so, the processes and conditions that make this approach especially effective.

3. Motivation to Teach and to Learn PA

Research has focused on the cognitive and linguistic factors involved in teaching PA to children. However, if teachers are not motivated to teach this skill, or if children are not motivated to learn it, then attention to it may be slighted. Some forms of teaching and learning are interesting and fun whereas other forms are tedious and boring. Research is needed to assess motivational properties of PA training programs and ways of enhancing motivation and interest if they are lacking.

4. Teaching PA With Computers

Use of computers is fast becoming a national pastime 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 the NRP database examined whether computers could deliver PA instruction effectively. Findings showed that effect sizes were 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 more effectively using computers.

5. Programs to Help Parents Teach PA

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 reviewed utilized parents as trainers. Research is needed to address this gap in our knowledge. In addition to informal activities that parents might use to draw children's attention to sounds in words, the effectiveness of activities that help parents teach letters to preschoolers might be explored and assessed.

6. High-Quality Research

Results of the NRP meta-analysis reveal the value of experimental studies for providing reliable findings that can guide instructional practice. The Panel 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. It is important for future researchers to maintain the quality of the designs adopted. This is not to say that all studies must use random assignment rather than nonequivalent groups. Sometimes experimenters have no choice if they want to conduct studies in school classrooms. However, researchers must take steps to maximize the rigor of their studies by addressing as many threats to internal and external validity as possible. Not only does this enhance confidence in the findings but also, as the NRP meta-analysis shows, it gives researchers a better chance of detecting treatment effects when they exist.

References

Note: An asterisk marks those publications providing 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 Publishing Co.

Anderson, R., Hiebert, E., Scott, J. & Wilkinson, I. (1985). Becoming a nation of readers: The report of the commission on reading. Washington, DC: National Academy of Education, Commission on Education and Public Policy.

Apthorp, H. (1998, April). Phonological awareness training for beginning readers: A meta-analysis of reading posttests. Paper presented at the American Educational Research Association Annual Meeting, 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.

Balmuth, M. (1982). The roots of phonics: a historical introduction. New York: Teachers College Press.

* 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. (in press). Phonological awareness. In M. Kamil, P. Mosenthal, P. Pearson, & R. Barr (Eds.), Handbook of reading research (Vol. 3). 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.

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: The University of Michigan Press. (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.

* 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. (First followup 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. (second followup 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.

Chall, J. (1967). Learning to read: the great debate. New York: McGraw Hill.

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.

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. (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), Language awareness and learning to read (pp. 119-147). New York: Springer Verlag.

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.

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

Glass, G., Cahen, L., Smith, M., & Filby, N. (1982). School class size. Beverly Hills, CA: Sage.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. Journal of Experimental Psychology: Human Perception and Performance, 5, 674-691.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. Journal of Experimental Child Psychology, 42, 73-83.

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. (in press). 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.

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.

* 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: The MIT Press.

Marsh, G., Freidman, M., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. In G. Mackinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3, pp. 199-221). New York: Academic Press.

* McGuiness, D., McGuiness, 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: The MIT Press.

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

* 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. (1998). 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. (followup to O'Connor et al., 1996)

* 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. (followup to Olofsson & Lundberg, 1983)

Pinnel, G., Lyons, C., DeFord, D., Bryk, A., & Seltzer M. (1994). Comparing instructional models for the literacy education of high-risk 1st 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.

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: The 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.

* 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., 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 and 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. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. Journal of Experimental Child Psychology, 39, 161-181.

* 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.

Tunmer, W., & Chapman, J. (1998). Language prediction skill, phonological recoding ability, and beginning reading. In C. Hulme and M. Joshi (Eds.), Reading and spelling: Development and disorders (pp. 33-68). Mahwah, NJ: Erlbaum.

* 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. (1997). 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. (1997). 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., King, J., & Olson, R. (1999). Training phonological awareness with and without explicit attention to articulation. Journal of Experimental Child Psychology, 72, 271-304.

* Wise, B., King, J., & Olson, R. (in press). Individual differences in benefits from computer-assisted remedial reading. Journal of Experimental Child Psychology.

Appendix A

Studies Included in the Meta-Analysis

* 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., 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.

* 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: The University of Michigan Press. (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.

* 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.

* 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. (First followup 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. (Second followup 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.

* 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.

* 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.

* 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.

* 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.

* 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.

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

* 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.

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

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

* 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. (1998). 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. (followup to O'Connor et al., 1996)

* 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. (Followup to Olofsson & Lundberg, 1983)

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

* 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.

* 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.

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

* 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.

* 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. (1997). Community-based early reading intervention for at-risk first graders. Learning Disabilities Research and Practice, 12, 29-39.

* Vadasy, P., Jenkins, J., Antil, L., et al. (1997). 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.

* Warrick, N., Rubin, H., & Rowe-Walsh, S. (1993). Phoneme awareness in language-delayed children: Comparative studies and intervention. Annals of Dyslexia, 43, 153-173.

* 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., King, J., & Olson, R. (1999). Training phonological awareness with and without explicit attention to articulation. Journal of Experimental Child Psychology, 72, 271-304.

* Wise, B., King, J., & Olson, R. (in press). Individual differences in benefits from computer-assisted remedial reading. Journal of Experimental Child Psychology.

Appendix B

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 vs. experimenter-devised tests of word reading
   - Standardized vs. experimenter-devised tests of nonword reading
   - Reading comprehension
4. Measures of spelling
   - Standardized vs. experimenter-devised tests of spelling
5. Measure of math achievement
6. Test points
   - Immediately after training
   - First followup test (delay of 2 to 15 months)
   - Second followup test (delay of 7 to 36 months)

PROPERTIES OF PHONEMIC AWARENESS TRAINING

1. PA skills taught:
   - a. Single skill; 2 skills; 3 or more skills
   - b. Segmenting and blending vs. 3 or more skills
2. Use of letters: phonemes and letters manipulated vs. only phonemes manipulated
3. Training unit: individuals; small groups (2 to 7 students); classrooms
4. Identity of trainer: classroom teachers; computers; researchers/others
5. Length of training: ranged from 1 hour to 75 hours

CHARACTERISTICS OF PARTICIPANTS

1. Reader level: at-risk readers; disabled readers; normally progressing readers
2. Grade level: preschool; kindergarten; 1st grade; 2nd through 6th grades
3. Language: English; other (Dutch, Finnish, German, Hebrew, Norwegian, Spanish, Swedish)
4. Socioeconomic status: low SES; middle-to-high SES

Table 1 (continued)

FEATURES OF THE DESIGN

1. Group assignment: random; matched; non-equivalent
2. Fidelity of trainers checked vs. 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)

Appendix C

Table 2: Phonemic Awareness Outcomes

Phonemic Awareness Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Time of Posttest					
Immediate	72	0.86*	No	0.79 to 0.92	ns
Followup	14	0.73*	Yes	0.61 to 0.85	
Outcome Measures of PA					
Segmentation (S)	51	0.87*	No	0.79 to 0.94	S = D > B
Blending (B)	33	0.61*	No	0.52 to 0.69	S > O
Deletion (D)	25	0.82*	No	0.73 to 0.91	B = O
Other (O)	37	0.72*	No	0.64 to 0.81	D = O
Characteristics of PA Training					
1 skill taught (1)	18	1.16*	No	0.96 to 1.36	1 = 2 > 3 +
2 skills (2)	24	1.03*	No	0.92 to 1.14	
3 or more skills (3)	30	0.70*	No	0.61 to 0.78	
Blend & segment only	18	0.81*	No	0.67 to 0.95	ns
3 or more skills	30	0.70*	No	0.61 to 0.78	
Letters manipulated	39	0.89*	No	0.80 to 0.98	ns
Letters not manipulated.	33	0.82*	No	0.73 to 0.91	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Individual child (I)	24	0.60*	Yes	0.47 to 0.72	S > I = C
Small groups (S)	35	1.38*	No	1.26 to 1.50	
Classrooms (C)	13	0.67*	No	0.57 to 0.76	
Length of training					
1 to 4.5 hrs (1)	15	0.61*	Yes	0.41 to 0.81	2 = 3 > 1 = 4
5 to 9.3 hrs (2)	24	1.37*	No	1.23 to 1.51	
10 to 18 hrs (3)	9	1.14*	No	0.97 to 1.32	
20 to 75 hrs (4)	22	0.65*	No	0.56 to 0.74	
Characteristics of Trainers					
Classroom teachers (CT)	19	0.78*	No	0.70 to 0.87	RO > CT
Researchers & others (RO)	53	0.94*	No	0.84 to 1.03	
Computers (Com)	8	0.66*	Yes	0.52 to 0.85	O > Com
Others (O)	64	0.89*	No	0.82 to 0.96	
Characteristics of Participants					
Reading level					
At risk (A)	15	0.95*	No	0.76 to 1.14	A = N > D
Disabled (D)	15	0.62*	No	0.48 to 0.75	
Normal progress (N)	42	0.93*	No	0.85 to 1.01	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Grade					
Preschool (Pre)	2	2.37*	No	1.93 to 2.81	Pre > K > 1 = 2
Kindergarten (K)	39	0.95*	No	0.87 to 1.04	
1st (1)	15	0.48*	Yes	0.31 to 0.64	
2nd-6th (2)	16	0.70*	Yes	0.56 to 0.83	
Socioeconomic status					
Low	12	1.07*	No	0.93 to 1.20	ns
Mid & High	17	1.02*	No	0.87 to 1.18	
Language					
English (E)	61	0.99*	No	0.90 to 1.07	E > O
Other (O)	11	0.65*	Yes	0.55 to 0.76	
Characteristics of Design					
Random assignment	33	0.87*	No	0.77 to 0.97	ns
Matched	18	0.92*	No	0.75 to 1.09	
Non-equivalent	21	0.83*	No	0.73 to 0.92	
Fidelity checked (FCh)	29	0.66*	No	0.56 to 0.75	Not > FCh
Not checked (Not)	43	1.02*	No	0.93 to 1.11	
Treated controls	38	0.89*	No	0.79 to 0.99	ns
Untreated controls	34	0.83*	No	0.75 to 0.92	

Table 2 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Size of sample					
9 to 22 students (1)	15	1.37*	No	1.09 to 1.66	1 = 3 > 2 = 4
24 to 30 students (2)	22	0.70*	No	0.53 to 0.87	
31 to 53 students (3)	13	1.10*	No	0.90 to 1.30	
56 to 383 students (4)	22	0.82*	No	0.74 to 0.89	
Characteristics of Study					
Year of publication					
1976-1985 (1)	10	0.73*	Yes	0.53 to 0.94	3 > 1 = 2 = 4
1986-1990 (2)	16	0.72*	No	0.59 to 0.85	
1991-1995 (3)	31	1.18*	No	1.07 to 1.30	
1996-2000 (4)	15	0.70*	No	0.59 to 0.81	

* indicates that effect size was significantly greater than zero at $p < 0.05$. ns indicates not significantly different from zero.

Table 3: Reading Outcomes

Reading Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Outcome Measures					
Time of posttest					
Immediate (Im)	90	0.53*	No	0.47 to 0.58	Im = 1 > 2
1st followup (1)	35	0.45*	No	0.36 to 0.54	
2nd followup (2)	8	0.23*	No	0.11 to 0.34	
Type of word test					
Experimenter (E)	58	0.61*	No	0.54 to 0.69	E > S
Standardized (S)	39	0.33*	No	0.24 to 0.42	
Type of pseudoword test					
Experimenter	47	0.56*	No	0.48 to 0.64	ns
Standardized	8	0.49*	Yes	0.29 to 0.69	
Reading comprehension	18	0.32*	No	0.18 to 0.46	
Math achievement	15	0.03ns	No	-0.11 to 0.16	
Characteristics of PA Training					
Immediate posttest					
1 skill taught (1)	32	0.71*	No	0.58 to 0.84	1 = 2 > 3 +
2 skills taught (2)	29	0.79*	No	0.69 to 0.89	
3 or more skills (3)	29	0.27*	Yes	0.19 to 0.35	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Followup posttest					
1 skill taught (1)	11	0.55*	Yes	0.37 to 0.73	2 > 1 > 3 +
2 skills (2)	9	1.28*	No	0.56 to 0.89	
3 or more skills (3)	15	0.23*	Yes	0.11 to 0.37	
Blend & segment only (BS)	19	0.67*	No	0.54 to 0.81	BS > 3 +
3 or more skills (3)	29	0.27*	Yes	0.19 to 0.35	
Immediate posttest					
Letters manipulated (L)	48	0.67*	No	0.59 to 0.75	L > NoL
Letters not manipulated (NoL)	42	0.38*	No	0.30 to 0.46	
Followup posttest					
Letters manipulated (L)	16	0.59*	No	0.45 to 0.74	L > NoL
Letters not manipulated (NoL)	19	0.36*	No	0.25 to 0.47	
Immediate posttest					
Individual child (I)	32	0.45*	Yes	0.34 to 0.57	S > I = C
Small groups (S)	42	0.81*	No	0.71 to 0.92	
Classrooms (C)	16	0.35*	No	0.26 to 0.44	
Followup posttest					
Individual child (I)	7	0.33*	Yes	0.11 to 0.55	S > I = C

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Small groups (S)	18	0.83*	No	0.66 to 1.00	
Classrooms (C)	10	0.30*	Yes	0.18 to 0.42	
Length of training					
1 to 4.5 hrs (1)	17	0.61*	Yes	0.42 to 0.79	1 = 2 = 3 > 4
5 to 9.3 hrs (2)	23	0.76*	No	0.62 to 0.89	
10 to 18 hrs (3)	19	0.86*	No	0.72 to 1.00	
20 to 75 hrs (4)	25	0.31*	No	0.22 to 0.39	
Characteristics of Trainers					
Immediate posttest					
Classroom teachers (CT)	22	0.41*	No	0.33 to 0.49	RO > CT
Researchers & others (RO)	68	0.64*	No	0.56 to 0.73	
Followup posttest					
Classroom teachers (CT)	12	0.32*	Yes	0.20 to 0.43	RO > CT
Researchers & others (RO)	23	0.63*	No	0.49 to 0.77	
Computers (Com)	8	0.33*	Yes	0.16 to 0.49	O > Com
Others (O)	82	0.55*	No	0.49 to 0.61	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Participants					
Reading level: Immediate posttest					
At risk (A)	27	0.86*	No	0.72 to 1.00	A > D = N
Disabled (D)	17	0.45*	Yes	0.32 to 0.57	
Normal progress (N)	46	0.47*	No	0.39 to 0.54	
Reading level: Followup posttest					
At risk	15	1.33*	No	1.10 to 1.56	A > D = N
Disabled	8	0.28*	Yes	0.10 to 0.46	
Normal progress	12	0.30*	Yes	0.19 to 0.42	
Grade					
Preschool (Pre)	7	1.25*	No	1.01 to 1.50	Pre > K = 1 = 2
Kindergarten (K)	40	0.48*	No	0.40 to 0.56	
1st (1)	25	0.49*	Yes	0.36 to 0.62	
2nd-6th (2)	18	0.49*	Yes	0.35 to 0.62	
Socioeconomic status					
Low (L)	11	0.45*	No	0.33 to 0.58	MH > L
Mid & High (MH)	29	0.84*	No	0.72 to 0.96	
Language					
Immediate posttest					
English (E)	72	0.63*	No	0.55 to 0.70	E > O
Other (O)	18	0.36*	No	0.27 to 0.46	

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Followup posttest					
English (E)	17	0.42*	Yes	0.28 to 0.56	ns
Other (O)	18	0.47*	No	0.35 to 0.59	
Characteristics of Design					
Random assignment (R)	46	0.63*	No	0.54 to 0.72	R > N
Matched (M)	22	0.57*	Yes	0.43 to 0.72	M = all
Nonequivalent (N)	20	0.40*	No	0.31 to 0.49	
Fidelity checked (FCh)	31	0.43*	No	0.34 to 0.53	Not > FCh
Not checked (Not)	59	0.59*	No	0.51 to 0.66	
Immediate posttest					
Treated controls (T)	54	0.65*	No	0.56 to 0.73	T > U
Untreated controls (U)	36	0.41*	No	0.33 to 0.49	
Followup posttest					
Treated contols (T)	20	0.62*	No	0.48 to 0.75	T > U
Untreated controls (U)	15	0.32*	Yes	0.20 to 0.44	
Size of sample					
9 to 22 students (1)	24	0.72*	No	0.51 to 0.92	1 = 3 > 4
24 to 30 students (2)	22	0.54*	Yes	0.37 to 0.70	2 = 1, 4
31 to 53 students (3)	22	0.91*	No	0.76 to 1.05	3 > 2

Table 3 (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
56 to 383 students (4)	22	0.40*	No	0.33 to 0.48	
Characteristics of Study					
Year of publication					
1976-1985 (1)	20	0.77*	No	0.62 to 0.93	1 = 3 > 2 = 4
1986-1990 (2)	16	0.36*	Yes	0.24 to 0.49	
1991-1995 (3)	41	0.77*	No	0.67 to 0.87	
1996-2000 (4)	13	0.21*	Yes	0.11 to 0.32	

* indicates that effect size was significantly greater than zero at $p < 0.05$.

ns indicates not significantly different from zero.

Table 4: Spelling Outcomes

Spelling Outcomes: Mean Effect Sizes (d) as a Function of Moderator Variables and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$. Effect Sizes Are Immediately After Training Unless Labeled as Followup.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Characteristics of Outcome Measures					
Time of Posttest					
Immediate (Im)	39	0.59*	No	0.49 to 0.68	Im > 1 = 2
1st followup (1)	17	0.37*	Yes	0.26 to 0.48	
2nd followup (2)	6	0.20*	No	0.08 to 0.32	
Type of spelling test					
Experimenter (E)	24	0.75*	No	0.62 to 0.89	E > S
Standardized (S)	20	0.41*	No	0.29 to 0.53	
Characteristics of PA Training					
1 skill taught (1)	17	0.74*	No	0.56 to 0.92	1 = 2 > 3 +
2 skills (2)	12	0.87*	Yes	0.71 to 1.03	
3 or more skills (3)	10	0.23*	No	0.07 to 0.38	
Blend & segment only (BS)	7	0.79*	Yes	0.49 to 1.09	BS > 3 +
3 or more skills (3)	10	0.23*	No	0.07 to 0.38	
Letters manipulated (L)	27	0.61*	No	0.50 to 0.72	L > NoL
Letters not used (NoL)	12	0.34*	No	0.25 to 0.42	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Individual child (I)	14	0.36*	No	0.20 to 0.52	S > I
Small groups (S)	20	0.77*	No	0.63 to 0.90	C = all
Classrooms (C)	5	0.56*	No	0.33 to 0.78	
Length of training					
1 to 4.5 hrs (1)	0	Ñ	Ñ	Ñ	
5 to 9.3 hrs (2)	8	1.13*	Yes	0.86 to 1.39	2 = 3 > 4
10 to 18 hrs (3)	10	0.87*	No	0.69 to 1.05	
20 to 75 hrs (4)	18	0.32*	No	0.19 to 0.45	
Characteristics of Trainers					
Classroom teachers (CT)	9	0.74*	No	0.58 to 0.90	CT > RO
Researchers & others (RO)	30	0.51*	No	0.39 to 0.62	
Computers (Com)	6	0.09ns	Yes -0.10 to 0.28	O > Com	
Others (O)	33	0.74*	No	0.63 to 0.85	
Characteristics of Participants					
Reading level					
At risk (A)	13	0.76*	No	0.54 to 0.98	A = N > D
Disabled (D)	11	0.15ns	Yes	-0.00 to 0.31	
Normal progress (N)	15	0.88*	No	0.74 to 1.02	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Grade					
Preschool (P)	0	Ñ	Ñ	Ñ	
Kindergarten (K)	15	0.97*	No	0.82 to 1.13	K > 1 > 2
1st (1)	16	0.52*	No	0.37 to 0.68	
2nd-6th (2)	8	0.14ns	Yes	-0.04 to 0.33	
Socioeconomic status					
Low (L)	6	0.76*	Yes	0.57 to 0.95	MH > L
Mid & High (MH)	9	1.17*	No	0.88 to 1.47	
Language					
English	32	0.60*	No	0.49 to 0.70	ns
Other	7	0.55*	Yes	0.31 to 0.78	
Characteristics of Design					
Random assignment (R)	17	0.37*	No	0.23 to 0.50	M = N > R
Matched (M)	12	0.73*	No	0.52 to 0.93	
Nonequivalent (N)	10	0.86*	Yes	0.69 to 1.04	
Fidelity checked (FCh)	15	0.44*	No	0.30 to 0.59	Not > FCh
Not checked (Not)	24	0.69*	No	0.57 to 0.81	

Table 4: Spelling Outcomes (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Treated controls (T)	24	0.43*	No	0.30 to 0.55	U > T
Untreated controls (U)	15	0.82*	No	0.67 to 0.96	
Size of sample					
9 to 22 students (1)	15	0.85*	Yes	0.59 to 1.10	2 > all
24 to 30 students (2)	3	1.68*	Yes	1.15 to 2.21	1 > 4
31 to 53 students (3)	8	0.75*	No	0.51 to 0.98	3 = 1, 4
56 to 383 students (4)	13	0.45*	No	0.34 to 0.56	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero

Appendix D

Table 5: Results

Mean Effect Sizes (d) With Reading Disabled Comparisons Removed from the Data Base and Tests to Determine Whether Effect Sizes Were Significantly Greater Than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
SPELLING OUTCOMES					
PA Skills Taught					
1 skill taught	14	0.77*	No	0.58 to 0.96	ns
2 skills taught	11	0.89*	Yes	0.72 to 1.05	
3 or more skills	3	0.93*	No	0.52 to 1.33	
Blend & segment only	6	0.85*	Yes	0.54 to 1.16	ns
3 or more skills	3	0.93*	No	0.52 to 1.33	
Letters manipulated (L)	17	1.00*	Yes	0.85 to 1.15	L > NoL
Letters not manipulated (NoL)	11	0.57*	No	0.37 to 0.76	
Training Unit					
Individual child (I)	8	1.00*	No	0.71 to 1.28	I = S > C
Small groups (S)	15	0.94*	Yes	0.78 to 1.10	
Classrooms (C)	5	0.56*	No	0.33 to 0.78	
Length of training					
1 to 4.5 hrs	0	0	Ñ	Ñ	
5 to 9.3 hrs	8	1.13*	Yes	0.86 to 1.39	ns
10 to 18 hrs	8	0.91*	No	0.73 to 1.10	
20 to 75 hrs	9	0.75*	Yes	0.50 to 1.01	

Table 5: Results (continued)

Moderator Variables and Levels	No. Cases	Mean d	Homogen.	95% CI	Contrasts
Trainer					
Classroom teachers	8	0.74*	No	0.58 to 0.91	ns
Researchers and others	20	0.96*	No	0.79 to 1.14	
Grade					
Preschool (Pre)	0				
Kindergarten (K)	15	0.97*	No	0.82 to 1.13	K > 1
1st (1)	13	0.66*	No	0.48 to 0.85	
2nd-6th (2)	0				
Language					
English (E)	22	0.95*	No	0.82 to 1.09	E > O
Other (O)	6	0.51*	Yes	0.28 to 0.75	
PHONEMIC AWARENESS OUTCOMES					
Letters manipulated (L)	25	1.11*	No	0.99 to 1.23	L > NoL
Letter not manipulated (NoL)	32	0.83*	No	0.73 to 0.92	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Appendix E

Table 6

Phonemic Awareness Outcomes: Mean Effect Sizes (d) Asssociated With Troia's Indicators of Methodological Rigor and Tests to Determine Whether Effect Sizes Were Significantly Greater than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Internal Validity				
% of criteria met				
24-40% (1)	10	0.67*	Yes	2 = 4 > 1
47% (2)	5	1.35*	No	4 > 3
53% (3)	14	0.95*	No	2 = 3
59-82% (4)	14	1.66*	No	
Critical Flaws				
1-2 (1)	18	1.63*	No	1 > 3 > 2
3 (2)	14	0.57*	Yes	
4-5 (3)	11	0.97*	No	
External Validity				
% of criteria met				
47-53% (1)	10	0.92*	No	4 > 1 = 2
56-60% (2)	14	0.81*	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	0.96*	No	1 = 2 = 3
2	13	0.61*	Yes	
3	9	0.97*	No	

Table 6 (continued)

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Ranking				
High rigor (1-12) (1)	15	1.56*	No	1 = 2 > 3
Mid (13-24) (2)	11	1.40*	No	
Low (25-36) (3)	17	0.69*	Yes	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Table 7

Reading Outcomes: Mean Effect Sizes (d) Asssociated With Troia's Indicators of Methodological Rigor and Tests to Determine Whether Effect Sizes Were Significantly Greater than Zero at $p < 0.05$, Were Homogeneous at $p < 0.05$, and Differed From Each Other at $p < 0.05$.

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Internal Validity				
% of criteria met				
24-40% (1)	11	0.49*	No	2 > 1
47% (2)	15	0.85*	No	4 > 1
53% (3)	16	0.63*	No	2 = 3 = 4
59-82% (4)	14	0.83*	No	1 = 3
Critical Flaws				
1-2 (1)	22	0.99*	No	1 > 2 = 3
3 (2)	18	0.59*	Yes	
4-5 (3)	16	0.56*	No	
External Validity				
% of criteria met				
47-53% (1)	16	0.98*	No	1 > 2, 3
56-60% (2)	14	0.58*	Yes	1 = 4
63-67% (3)	15	0.61*	No	2 = 3 = 4
73-81% (4)	11	0.66*	No	

Table 7 (continued)

Variables and Levels	No. Cases	Mean d	Homogen.	Contrasts
Critical Flaws				
0 flaws	17	0.90*	No	0 = 3 > 1
1	11	0.51*	No	2 = all
2	17	0.57*	Yes	
3	11	0.92*	No	
Ranking				
High rigor (1-12) (1)	19	1.00*	No	1 > 2 = 3
Mid (13-24) (2)	14	0.61*	Yes	
Low (25-36) (3)	23	0.58*	No	

* indicates that effect size was significantly greater than zero at $p < 0.05$.
ns indicates not significantly different from zero.

Appendix F

Studies in the Phonemic Awareness (PA) Database, Their Characteristics, and Effect Sizes

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
Ball & Blachman, 1991	.	.	.	.	.	.	.	.	.	.	.	.	89	.	.	.
01 - Segment & categ. + let vs. Language, LS	2	Yes	SmG	Other	9.33	Nor	K	Engl	.	R	Yes	59	.	1.49	0.71	0.87
02 - Segment & categ. + 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 & categ. vs. Language	2	No	SmG	Other	10	AR	K	Hebr	M-H	R	No	50	.	.	4.21	.
05 - Segment & categ. vs. No treatment	2	No	SmG	Other	10	AR	K	Hebr	M-H	R	No	41	.	.	4.33	.
06 - Segment & categ. + let vs. Language	2	Yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	50	.	.	2.1	.
07 - Segment & categ. + let vs. No treat.	2	Yes	SmG	Other	10	AR	K	Hebr	M-H	R	No	41	.	.	2.17	.
Blachman et al., 1994	.	.	.	.	.	.	.	.	.	.	.	.	159	.	.	.
08 - Segment & categ. + let vs. No treat.	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 - Phon. categ. vs. Semantic categ.	1	No	Ind	Other	11.67	AR	1st	Engl	.	M/R	No	39		.	0.5	0.39
10 - Phon. categ. vs. No treatment	1	No	Ind	Other	11.67	AR	1st	Engl	.	M/R	No	26	.	.	0.86	1
11 - Phon. categ. + let vs. Semantic categ.	1	Yes	Ind	Other	11.67	AR	1st	Engl	.	M/R	No	39	.	.	1.17	1.59
12 - Phon. categ. + let vs. No treatment	1	Yes	Ind	Other	11.67	AR	1st	Engl	.	M/R	No	26	.	.	1.53	2.18

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
Brady et al., 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. Pre-read prep., LS	2	No	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	130	.	0.55	0.54	.
16 - Segment & blend + let vs. Pre-read prep., LS	2	Yes	Clas	Teach	5	Nor	K	Dutch	M-H	R	Yes	134	.	0.25	0.35	.
Byrne & Fielding-Barnsley, 1991, '93, '95	.	.	.	.	.	.	.	.	.	.	.	.	126	.	.	.
17 - Phon. categ. + let vs. Semantic categ.	1	Yes	SmG	Other	6	Nor	Pre	Engl	M-H	R	No	126	.	3.14	1.61	.34*
Castle, et al., 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, meta. 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, meta. vs. Stories	2	No	SmG	Other	6	Nor	1st	Engl	M-H	M/R	No	28	.	1.27	0.51	.
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

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
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 - Categ. + let vs. Semantic categ.	1	Yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22	.	.	0.82	1.44
28 - Categ. + let vs. Hand manipulation	1	Yes	SmG	Other	30	AR	1st	Span	M-H	R	No	22	.	.	0.73	1.03
29 - Categ. vs. Semantic categ.	1	No	SmG	Other	30	AR	1st	Span	M-H	R	No	21	.	.	0.18	0.36
30 - Categ. vs. Hand manipulation	1	No	SmG	Other	30	AR	1st	Span	M-H	R	No	21	.	.	0.14	0.02
Ehri & Wilce, 1987	.	.	.	.	.	.	.	.	.	.	.	.	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 et al., 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 treat., LS	1	No	SmG	Other	5	AR	K	Engl	.	R	No	21	.	0.75	-0.19	.
37 - Onset-rime, LS vs. No treat., LS	1	No	SmG	Other	5	AR	K	Engl	.	R	No	21	.	1.6	3.6	.

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
Gross & Garnet, 1994	.	.	.	.	.	.	.	.	.	.	.	.	12	.	.	.
38 - Categ. 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 et al., 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 Rec. vs. Read Rec.	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 treat.	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 Rec. vs. Read Rec.	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 & Peltoma, 1993	.	.	.	.	.	.	.	.	.	.	.	.	46	.	.	.
48 - Blend & categ. + let vs. Speech therapy	2	Yes	SmG	Other	.	AR	K	Fin	.	NE	No	46	.	.	.60*	.67*

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	No. skills	Letters	Tr.unit	Trainer	Length in hours	Reader	Grade	Language	SES	Group Assign.	Fidelity	N (Case)	N (Study)	PA	Read	Spell
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 - Categ. 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
Lovett et al., 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 et al., 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 et al., 1995, Study 2	.	.	.	.	.	.	.	.	.	.	.	.	42	.	.	.
55 - Mult. PA + let in Montessori vs. No treat.	3+	Yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27	.	0.15	1.11	.
56 - Mult. PA + let in whole lang. vs. No treat.	3+	Yes	SmG	Teach	66.67	Nor	1st	Engl	M-H	NE	Yes	27	.	0.37	1.22	.
Murray, 1998	.	.	.	.	.	.	.	.	.	.	.	.	48	.	.	.
57 - Categ., 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 et al., 1995	.	.	.	.	.	.	.	.	.	.	.	.	67	.	.	.
60 - Segment & blend, LS vs. LS	2	No	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45	.	2.69	1.64	.

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
61 - Mult. PA, LS vs. LS	3+	No	SmG	Other	5	AR	K	Engl	Lo	R	Yes	45	.	2.42	0.52	.
O'Connor et al., 1996, 1998	.	.	.	.	.	.	.	.	.	.	.	.	80	.	.	.
62 - Segment & blend + let vs. No treat.	2	Yes	Clas	Teach	20	Nor	K	Engl	.	NE	Yes	66	.	0.62	0.11	0.73
63 - Segment & blend + let vs. No treat.	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. Vocab. comput.	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 et al., 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 & categ. + let vs. No treat.	2	Yes	SmG	Teach	13.2	Nor	K	Engl	Lo	NE	No	149	.	1.81	0.67	0.94

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
Torgesen et al., 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	.
Treiman & Baron, 1983	.	.	.	.	.	.	.	.	.	.	.	.	28	.	.	.
75 - Onset-rime vs. Repeat syllables	1	No	Ind	Other	.	Nor	Pre	Engl	M-H		Yes	8	.	.	0.62	.
76 - Onset-rime vs. Repeat syllables	1	No	Ind	Other	.	Nor	K	Engl	M-H		Yes	20	.	.	0.13	.
Uhry & Shepherd, 1993	.	.	.	.	.	.	.	.	.	.	.	.	22	.	.	.
77 - Seg. & blend + let vs. Text	2	Yes	SmG	Comp	17.33	Nor	1st	Engl	M-H	R	No	22	.	1.45	1.07	0.77
Vadasy et al., 1997 (LDRP)	.	.	.	.	.	.	.	.	.	.	.	.	35	.	.	.
78 - Segment & categ. + let vs. No treat.	2	Yes	Ind	Other	54	AR	1st	Engl	Lo	R	Yes	35	.	0.74	0.44	0.67
Vadasy et al., 1997 (LDQ)	.	.	.	.	.	.	.	.	.	.	.	.	40	.	.	.
79 - Segment & blend + let vs. No treat.	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	0.3	.
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	0.33	.
87 - Mult. PA, let, word vs. Word	3+	Yes	Ind	Other	2.5	Nor	2nd+	Engl	.	R	No	30	.	0.66	0.52	.

Appendix F (continued)

	Characteristics of Training					Characteristics of Participants				Features of Design				Effect Sizes		
Author and Year, Treatment vs. Control	**No. skills**	**Letters**	**Tr.unit**	**Trainer**	**Length in hours**	**Reader**	**Grade**	**Language**	**SES**	**Group Assign.**	**Fidelity**	**N (Case)**	**N (Study)**	**PA**	**Read**	**Spell**
Warrick et al., 1993, Study II	.	.	.	.	.	.	.	.	.	.	.	.	28	.	.	.
88 - Segment vs. No treatment	1	No	SmG	Other	5.33	AR	K	Engl	.	NE	No	28	.	0.67	1.30*	.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 treat.	2	Yes	Clas	Teach	62.83	RD	2nd+	Engl	.	NE	Yes	102	.	0.35	1.05	.
92 - Segment & blend + let vs. No treat.	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	26.67	RD	2nd+	Engl	Lo	NE	Yes	48	.	0.12	0.47	0.49
Wise et al., 1999	.	.	.	.	.	.	.	.	.	.	.	.	122	.	.	.
94 - Mult. PA wi artic. + 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. Artic., LS	3+	Yes	SmG	Comp	42	RD	2nd+	Engl	.	R	Yes	85	.	0.66	0.28	0.3
Wise et al., 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
Abreviations:																

LS = Letter-sound training provided separately	**Mult**. = Multiple PA in 3 or more skills
Categ. = Categorization or identity training	**Recip. Teach** = Reciprocal teaching strategies learned and applied to reading
Meta. = Metacognitive activities to understand purposes, use of PA	***** = Effect sizes were drawn from followup test points.
Read Rec. = Reading Recovery® program	