National Reading Panel

TEACHING CHILDREN TO READ:
An Evidence-Based Assessment
of the Scientific Research Literature
on Reading and Its Implications
for Reading Instruction

Acknowledgments

The National Reading Panel wishes to express its gratitude to the following individuals for their contributions to its effort.

Marilyn Adams
Michelle Eidlitz
Blair Johnson
Khalil Nourani
Michael Pressley
Robin Sidhu
Ed Bouchard
Barbara Foorman
Alisa Kenny
Simone Nunes
David Reinking
Steven Stahl
Harris Cooper
David Francis
Helen S. Kim
Elizabeth S. Pang
Scott J. Ross
Maggie Toplak
Gerald Duffy
Ester Halberstam
Marjolaine Limbos
Joan Pagnucco
Barbara Schuster
Zoreh Yaghoubzadeh

Members of the National Reading Panel

Donald N. Langenberg, Ph.D., *Chair*

Gloria Correro, Ed.D.
Gwenette Ferguson, M.Ed.
Michael L. Kamil, Ph.D.
S.J. Samuels, Ed.D.
Sally E. Shaywitz, M.D.
Joanna Williams, Ph.D.
Joanne Yatvin, Ph.D.
Linnea Ehri, Ph.D.
Norma Garza, C.P.A.
Cora Bagley Marrett, Ph.D.
Timothy Shanahan, Ph.D.
Thomas Trabasso, Ph.D.
Dale Willows, Ph.D.

Members of the National Reading Panel Subgroups

Alphabetics
Linnea Ehri, *Chair*
Gloria Correro
Timothy Shanahan
Dale Willows
Joanne Yatvin

Comprehension
Michael L. Kamil, *Chair*
Gwenette Ferguson
Norma Garza
Thomas Trabasso
Joanna Williams

Fluency
S.J. Samuels, *Co-Chair*
Timothy Shanahan, *Co-Chair*
Sally E. Shaywitz

Methodology
Timothy Shanahan, *Co-Chair*
Sally E. Shaywitz, *Co-Chair*

Teacher Education
Gloria Corerro, *Co-Chair*
Michael L. Kamil, *Co-Chair*
Gwenette Ferguson
Norma Garza
Cora Bagley Marrett

Technology/Next Steps
Michael L. Kamil, *Chair*
Donald N. Langenberg

Staff of the National Reading Panel

F. William Dommel, Jr., J.D., *Executive Director*

Vinita Chhabra, M.Ed., *Research Scientist*
Judith Rothenberg, *Secretary*
Jaimee Nusbacher, *Meeting Manager*
Mary E. McCarthy, Ph.D., *Senior Staff Psychologist*
Stephanne Player, *Support Staff*
Patrick Riccards, *Senior Advisor*

Table of Contents

(Report of the National Reading Panel: Reports of the Subgroups provides complete and extensive descriptions of information presented in this Report.)

Introduction

Congressional Charge

In 1997, Congress asked the "Director of the National Institute of Child Health and Human Development (NICHD), in consultation with the Secretary of Education, to convene a national panel to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read." This panel was charged with providing a report that "should present the panel's conclusions, an indication of the readiness for application in the classroom of the results of this research, and, if appropriate, a strategy for rapidly disseminating this information to facilitate effective reading instruction in the schools. If found warranted, the panel should also recommend a plan for additional research regarding early reading development and instruction."

Establishment of the National Reading Panel

In response to this Congressional request, the Director of NICHD, in consultation with the Secretary of Education, constituted and charged a National Reading Panel (the NRP or the Panel). The NRP comprised 14 individuals, including (as specified by Congress) "leading scientists in reading research, representatives of colleges of education, reading teachers, educational administrators, and parents." The original charge to the NRP asked that a final report be submitted by November 1998. When the Panel began its work, it quickly became apparent that the Panel could not respond properly to its charge within that time constraint. Permission was sought and received to postpone the report's submission deadline. A progress report was submitted to Congress in February 1999. The information provided in the NRP Progress Report, this Report of the National Reading Panel, and the Report of the National Reading Panel: Reports of the Subgroups reflect the findings and determinations of the National Reading Panel.

NRP Approach to Achieving the Objectives of Its Charge and Initial Topic Selection

The charge to the NRP took into account the foundational work of the National Research Council (NRC) Committee on Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998). The NRC report is a consensus document based on the best judgments of a diverse group of experts in reading research and reading instruction. The NRC Committee identified and summarized research literature relevant to the critical skills, environments, and early developmental interactions that are instrumental in the acquisition of beginning reading skills. The NRC Committee did not specifically address "how" critical reading skills are most effectively taught and what instructional methods, materials, and approaches are most beneficial for students of varying abilities.

In order to build upon and expand the work of the NRC Committee, the NRP first developed an objective research review methodology. The Panel then applied this methodology to undertake comprehensive, formal, evidence-based analyses of the experimental and quasi-experimental research literature relevant to a set of selected topics judged to be of central importance in teaching children to read. An examination of a variety of public databases by Panel staff revealed that approximately 100,000 research studies on reading have been published since 1966, with perhaps another 15,000 appearing before that time. Obviously, it was not possible for a panel of volunteers to examine critically this entire body of research literature. Selection of prioritized topics was necessitated by the large amount of published reading

research literature relevant to the Panel's charge to determine the effectiveness of reading instructional methods and approaches. A screening process was therefore essential.

The Panel's initial screening task involved selection of the set of topics to be addressed. Recognizing that this selection would require the use of informed judgment, the Panel chose to begin its work by broadening its understanding of reading issues through a thorough analysis of the findings of the NRC report, Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998). Early in its deliberations the Panel made a tentative decision to establish subgroups of its members and to assign to each of them one of the major topic areas designated by the NRC Committee as central to learning to read—Alphabetics, Fluency, and Comprehension.

Regional Public Hearings

As part of its information gathering, the Panel publicly announced, planned, and held regional hearings in Chicago, IL (May 29, 1998), Portland, OR (June 5, 1998), Houston, TX (June 8, 1998), New York, NY (June 23, 1998), and Jackson, MS (July 9, 1998). The Panel believed that it would not have been possible to accomplish the mandate of Congress without first hearing directly from consumers of this information—teachers, parents, students, and policymakers—about their needs and their understanding of the research. Although the regional hearings were not intended as a substitute for scientific research, the hearings gave the Panel an opportunity to listen to the voices of those who will need to consider implementation of the Panel's findings and determinations. The regional hearings gave members a clearer understanding of the issues important to the public.

As a result of these hearings, the Panel received oral and written testimony from approximately 125 individuals or organizations representing citizens—teachers, parents, students, university faculty, educational policy experts, and scientists—who would be the ultimate users and beneficiaries of the research-derived findings and determinations of the Panel.

At the regional hearings, several key themes were expressed repeatedly:

- The importance of the role of parents and other concerned individuals, especially in providing children with early language and literacy experiences that foster reading development;
- The importance of early identification and intervention for all children at risk for reading failure;
- The importance of phonemic awareness, phonics, and good literature in reading instruction and the need to develop a clear understanding of how best to integrate different reading approaches to enhance the effectiveness of instruction for all students;
- The need for clear, objective, and scientifically based information on the effectiveness of different types of reading instruction and the need to have such research inform policy and practice;
- The importance of applying the highest standards of scientific evidence to the research review process so that conclusions and determinations are based on findings obtained from experimental studies characterized by methodological rigor with demonstrated reliability, validity, replicability, and applicability;
- The importance of the role of teachers, their professional development, and their interactions and collaborations with researchers, which should be recognized and encouraged; and
- The importance of widely disseminating the information that is developed by the Panel.

Adoption of Topics To Be Studied

Following the regional hearings, the Panel considered, discussed, and debated several dozen possible topic areas and then settled on the following topics for intensive study:

- Alphabetics
  - Phonemic Awareness Instruction
  - Phonics Instruction

- Fluency
- Comprehension
  - Vocabulary Instruction
  - Text Comprehension Instruction
  - Teacher Preparation and Comprehension Strategies Instruction
- Teacher Education and Reading Instruction
- Computer Technology and Reading Instruction.

In addition, because of the concern voiced by the public at the regional hearings that the highest standards of scientific evidence be applied in the research review process, the methodology subgroup was tasked to develop a research review process including specific review criteria.

Each topic and subtopic became the subject of the work of a subgroup composed of one or more Panel members. Some Panel members served on more than one subgroup. The subgroups formulated seven broad questions to guide their efforts in meeting the Congressional charge of identifying effective instructional reading approaches and determining their readiness for application in the classroom:

1. Does instruction in phonemic awareness improve reading? If so, how is this instruction best provided?
2. Does phonics instruction improve reading achievement? If so, how is this instruction best provided?
3. Does guided oral reading instruction improve fluency and reading comprehension? If so, how is this instruction best provided?
4. Does vocabulary instruction improve reading achievement? If so, how is this instruction best provided?
5. Does comprehension strategy instruction improve reading? If so, how is this instruction best provided?
6. Do programs that increase the amount of children's independent reading improve reading achievement and motivation? If so, how is this instruction best provided?
7. Does teacher education influence how effective teachers are at teaching children to read? If so, how is this instruction best provided?

Each subgroup also generated several subordinate questions to address within each of the major questions. It should be made clear that the Panel did not consider these questions and the instructional issues that they represent to be the *only* topics of importance in learning to read. The Panel's silence on other topics should not be interpreted as indicating that other topics have no importance or that improvement in those areas would not lead to greater reading achievement. It was simply the sheer number of studies identified by Panel staff relevant to reading (more than 100,000 published since 1966 and more than 15,000 prior to 1966) that precluded an exhaustive analysis of the research in all areas of potential interest.

The Panel also did not address issues relevant to second language learning, as this topic was being addressed in detail in a new, comprehensive NICHD/ OERI (Office of Educational Research and Improvement) research initiative. The questions presented above bear on instructional topics of widespread interest in the field of reading education that have been articulated in a wide range of theories, research studies, instructional programs, curricula, assessments, and educational policies. The Panel elected to examine these and subordinate questions because they currently reflect the central issues in reading instruction and reading achievement. The methodological processes described in the next section guided the Panel's examination and analysis of the extant research.

Methodological Overview

In what may be its most important action, the Panel then developed and adopted a set of rigorous research methodological standards. (See the methodology adopted by the Panel and printed as an addendum to this Report.) These standards guided the screening of the research literature relevant to each topic area addressed by the Panel. This screening process identified a final set of experimental or quasi-experimental research studies that were then subjected to detailed analysis. The evidence-based methodological standards adopted by the Panel are essentially those normally used in research studies of the efficacy of interventions in psychological and medical research. These include behaviorally based interventions, medications, or medical procedures proposed for use in the fostering of robust health and psychological development and the prevention or treatment of disease.

It is the view of the Panel that the efficacy of materials and methodologies used in the teaching of reading and in the prevention or treatment of reading disabilities should be tested no less rigorously. However, such standards have not been universally accepted or used in reading education research. Unfortunately, only a small fraction of the total reading research literature met the Panel's standards for use in the topic analyses.

The research literature screening process proceeded essentially as follows. For each topic, an initial pool of candidate studies was created by searching a minimum of two databases (PsycINFO and ERIC) for study reports relevant to the topic. To be included in the database, studies had to measure reading as an outcome. Reading was defined to include several behaviors such as the following: reading real words in isolation or in context, reading pseudowords that can be pronounced but have no meaning, reading text aloud or silently, and comprehending text that is read silently or orally. From the pool produced by the electronic searches of the databases, those studies were selected that met the following criteria:

- Published in English in a refereed journal;
- Focused on children's reading development in the age/grade range from preschool to grade 12; and
- Used an experimental or quasi-experimental design with a control group or a multiple-baseline method.

Those studies meeting the above criteria formed the set of studies subjected to further analysis. The next step was to code each study for several characteristics including the following:

- Characteristics of study participants (age; demographics; cognitive, academic, and behavioral characteristics);
- Study interventions, described in sufficient detail to allow for replicability, including how long the interventions lasted and how long the effects lasted;
- Study methods, with sufficient description to allow judgments about how instruction fidelity was insured; and
- Nature of the outcome measures and whether they were described fully.

For each study meeting the above criteria, relevant reported statistics were coded in a standardized format and analyzed. For several topics, the number of studies meeting criteria was sufficient to permit a formal statistical meta-analysis, including calculation of effect sizes. For others, a full meta-analysis could not be carried out. Where there were too few studies that satisfied the Panel's criteria to permit a meta-analysis, the Panel made a decision to conduct a more subjective qualitative analysis to provide the best possible information about an instructional reading approach or program.

With this information as background, this Report is organized into sections to provide an overview of the major findings and determinations achieved by the NRP in the areas of alphabetics (phonemic awareness instruction and phonics instruction), fluency, comprehension (vocabulary instruction, text comprehension instruction, and teacher preparation and comprehension strategies instruction), teacher education and reading instruction, computer technology and reading instruction, and next steps. This Report concludes with some reflections on the NRP process and products.

Alphabetics

Phonemic Awareness Instruction

Phonemes are the smallest units composing spoken language. For example, the words "go" and "she" each consist of two sounds or phonemes. Phonemes are different from letters that represent phonemes in the spellings of words. Instruction in phonemic awareness (PA) involves teaching children to focus on and manipulate phonemes in spoken syllables and words. PA instruction is frequently confused with phonics instruction, which entails teaching students how to use letter-sound relations to read or spell words. PA instruction qualifies as phonics instruction when it involves teaching children to blend or segment the sounds in words using letters. However, children may be taught to manipulate sounds in speech without any letters as well; this does not qualify as phonics instruction. PA is also frequently confused with auditory discrimination, which refers to the ability to recognize whether two spoken words are the same or different. These distinctions are explained in detail in the section devoted to phonemic awareness instruction in the Report of the National Reading Panel: Reports of the Subgroups.

There are several reasons why the NRP selected PA instruction for review and analysis. First, correlational studies have identified PA 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. Such evidence suggests the potential importance of PA training in the development of reading skills. Second, many experimental studies have been carried out to evaluate the effectiveness of PA training in facilitating reading acquisition. Third, there is currently much interest in PA training programs among teachers, principals, parents, and publishers because of claims about their value in improving children's ability to learn to read.

The initial literature search for studies relevant to PA instruction and training identified 1,962 citations. Following initial review, the Panel identified and further reviewed 78 studies that met the general NRP research methodology criteria. However, on detailed examination, only 52 studies satisfied the more specific NRP research methodology criteria. From these 52 studies, 96 comparisons of treatment and control groups were derived. Data from these comparisons were then entered into a meta-analysis to determine treatment effect sizes.

Findings and Determinations

The results of the meta-analysis were impressive. Overall, the findings showed that teaching children to manipulate phonemes in words was highly effective under a variety of teaching conditions with a variety of learners across a range of grade and age levels and that teaching phonemic awareness to children significantly improves their reading more than instruction that lacks any attention to PA.

Specifically, the results of the experimental studies led the Panel to conclude that PA training was the cause of improvement in students' phonemic awareness, reading, and spelling following training. The findings were replicated repeatedly across multiple experiments and thus provide converging evidence for causal claims. While PA training exerted strong and significant effects on reading and spelling development, it did not have an impact on children's performance on math tests. This indicates that halo/Hawthorne (novelty) effects did not explain the findings and that indeed the training effects were directly connected with and limited to the targeted domain under study. Importantly, the effects of PA instruction on reading lasted well beyond the end of training. Children of varying abilities improved their PA and their reading skills as a function of PA training.

PA instruction also helped normally achieving children learn to spell, and the effects lasted well beyond the end of training. However, the instruction was not effective for improving spelling in disabled readers. This is consistent with other research showing that disabled readers have difficulty learning how to spell.

Programs in all of the studies provided explicit instruction in phonemic awareness. Specifically, the characteristics of PA training found to be most effective in enhancing PA, reading, and spelling skills included explicitly and systematically teaching children to manipulate phonemes with letters, focusing the instruction on one or two types of phoneme manipulations rather than multiple types, and teaching children in small groups.

PA instruction is ready for implementation in the classroom, but teachers should keep in mind several cautions. First, PA training does not constitute a complete reading program. Rather, it provides children with essential foundational knowledge in the alphabetic system. It is one necessary instructional component within a complete and integrated reading program. Several additional competencies must be acquired as well to ensure that children will learn to read and write. Second, there are many ways to teach PA effectively. In implementing PA instruction, teachers need to evaluate the methods they use against measured success in their own students. Third, the motivation of both students and their teachers is a critical ingredient of success. Research has not specifically focused on this.

Phonics Instruction

Phonics instruction is a way of teaching reading that stresses the acquisition of letter-sound correspondences and their use in reading and spelling. The primary focus of phonics instruction is to help beginning readers understand how letters are linked to sounds (phonemes) to form letter-sound correspondences and spelling patterns and to help them learn how to apply this knowledge in their reading. Phonics instruction may be provided systematically or incidentally. The hallmark of a systematic phonics approach or program is that a sequential set of phonics elements is delineated and these elements are taught along a dimension of explicitness depending on the type of phonics method employed. Conversely, with incidental phonics instruction, the teacher does not follow a planned sequence of phonics elements to guide instruction but highlights particular elements opportunistically when they appear in text.

Phonics Instructional Approaches

Analogy Phonics—Teaching students unfamiliar words by analogy to known words (e.g., 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 word onset, such as reading brick by recognizing that -ick is contained in the known word kick, or reading stump by analogy to jump).

Analytic Phonics—Teaching students to analyze letter-sound relations in previously learned words to avoid pronouncing sounds in isolation.

Embedded Phonics—Teaching students phonics skills by embedding phonics instruction in text reading, a more implicit approach that relies to some extent on incidental learning.

Phonics through Spelling—Teaching students to segment words into phonemes and to select letters for those phonemes (i.e., teaching students to spell words phonemically).

Synthetic Phonics—Teaching students explicitly to convert letters into sounds (phonemes) and then blend the sounds to form recognizable words.

Types of Phonics Instructional Methods and Approaches

The sidebar depicts several different types of phonics instructional approaches that vary according to the unit of analysis or how letter-sound combinations are represented to the student. For example, in synthetic phonics approaches, students are taught to link an individual letter or letter combination with its appropriate sound and then blend the sounds to form words. In analytic phonics, students are first taught whole word units followed by systematic instruction linking the specific letters in the word with their respective sounds. Phonics instruction can also vary with respect to the explicitness by which the phonic elements are taught and practiced in the reading of text. For example, many synthetic phonics approaches use direct instruction in teaching phonics components and provide opportunities for applying these skills in decodable text formats characterized by a controlled vocabulary. On the other hand, embedded phonics approaches are typically less explicit and use decodable text for practice less frequently, although the phonics concepts to be learned can still be presented systematically. These distinctions are addressed in detail in the Phonics subgroup report.

Questions Guiding the NRP Analysis of Phonics Instruction

The NRP examined the research literature concerning phonics instruction to answer the following questions: Does phonics instruction enhance children's success in learning to read? Is phonics instruction more effective at some grade levels than others? Is it beneficial for children who are having difficulties learning to read? Does phonics instruction improve all aspects of reading or just decoding and word-level reading skills? Are some types of phonics instruction more effective than others and for which children? Does phonics instruction have an impact on children's spelling?

To address these questions the NRP performed a literature search to identify studies published since 1970 that compared phonics instruction to other forms of instruction for their impact on reading ability. The initial electronic and manual searches identified 1,373 studies that appeared relevant to phonics instruction. Evaluation of these studies to determine adherence to the general and specific NRP research methodology criteria identified 38 studies from which 66 treatment-control group comparisons were derived. Data from these studies were used in a meta-analysis, including the calculation of effect sizes.

The meta-analysis indicated that systematic phonics instruction enhances children's success in learning to read and that systematic phonics instruction is significantly more effective than instruction that teaches little or no phonics.

Findings and Determinations

The meta-analysis revealed that systematic phonics instruction produces significant benefits for students in kindergarten through 6th grade and for children having difficulty learning to read. The ability to read and spell words was enhanced in kindergartners who received systematic beginning phonics instruction. First graders who were taught phonics systematically were better able to decode and spell, and they showed significant improvement in their ability to comprehend text. Older children receiving phonics instruction were better able to decode and spell words and to read text orally, but their comprehension of text was not significantly improved.

Systematic synthetic phonics instruction (see sidebar for definition) had a positive and significant effect on disabled readers' reading skills. These children improved substantially in their ability to read words and showed significant, albeit small, gains in their ability to process text as a result of systematic synthetic phonics instruction. This type of phonics instruction benefits both students with learning disabilities and low-achieving students who are not disabled. Moreover, systematic synthetic phonics instruction was significantly more effective in improving low socioeconomic status (SES) children's alphabetic knowledge and word reading skills than instructional approaches that were less focused on these initial reading skills.

Across all grade levels, systematic phonics instruction improved the ability of good readers to spell. The impact was strongest for kindergartners and decreased in later grades. For poor readers, the impact of phonics instruction on spelling was small, perhaps reflecting the consistent finding that disabled readers have trouble learning to spell.

Although conventional wisdom has suggested that kindergarten students might not be ready for phonics instruction, this assumption was not supported by the data. The effects of systematic early phonics instruction were significant and substantial in kindergarten and the 1st grade, indicating that systematic phonics programs should be implemented at those age and grade levels.

The NRP analysis indicated that systematic phonics instruction is ready for implementation in the classroom. Findings of the Panel regarding the effectiveness of explicit, systematic phonics instruction were derived from studies conducted in many classrooms with typical classroom teachers and typical American or English-speaking students from a variety of backgrounds and socioeconomic levels. Thus, the results of the analysis are indicative of what can be accomplished when explicit, systematic phonics programs are implemented in today's classrooms. Systematic phonics instruction has been used widely over a long period of time with positive results, and a variety of systematic phonics programs have proven effective with children of different ages, abilities, and socioeconomic backgrounds.

These facts and findings provide converging evidence that explicit, systematic phonics instruction is a valuable and essential part of a successful classroom reading program. However, there is a need to be cautious in giving a blanket endorsement of all kinds of phonics instruction.

It is important to recognize that the goals of phonics instruction are to provide children with key knowledge and skills and to ensure that they know how to apply that knowledge in their reading and writing. In other words, phonics teaching is a means to an end. To be able to make use of letter-sound information, children need phonemic awareness. That is, they need to be able to blend sounds together to decode words, and they need to break spoken words into their constituent sounds to write words. Programs that focus too much on the teaching of letter-sound relations and not enough on putting them to use are unlikely to be very effective. In implementing systematic phonics instruction, educators must keep the *end* in mind and ensure that children understand the purpose of learning letter sounds and that they are able to apply these skills accurately and fluently in their daily reading and writing activities.

Of additional concern is the often-heard call for "intensive, systematic" phonics instruction. Usually the term "intensive" is not defined. How much is required to be considered intensive? In addition, it is not clear how many months or years a phonics program should continue. If phonics has been systematically taught in kindergarten and 1st grade, should it continue to be emphasized in 2nd grade and beyond? How long should single instruction sessions last? How much ground should be covered in a program? Specifically, how many letter-sound relations should be taught, and how many different ways of using these relations to read and write words should be practiced for the benefits of phonics to be maximized? These questions remain for future research.

Another important area is the role of the teacher. Some phonics programs showing large effect sizes require teachers to follow a set of specific instructions provided by the publisher; while this may standardize the instructional sequence, it also may reduce teacher interest and motivation. Thus, one concern is how to maintain consistency of instruction while still encouraging the unique contributions of teachers. Other programs require a sophisticated knowledge of spelling, structural linguistics, or word etymology. In view of the evidence showing the effectiveness of systematic phonics instruction, it is important to ensure that the issue of how best to prepare teachers to carry out this teaching effectively and creatively is given high priority.

Knowing that all phonics programs are not the same brings with it the implication that teachers must themselves be educated about how to evaluate different programs to determine which ones are based on strong evidence and how they can most effectively use these programs in their own classrooms. It is therefore important that teachers be provided with evidence-based preservice training and ongoing inservice training to select (or develop) and implement the most appropriate phonics instruction effectively.

A common question with any instructional program is whether "one size fits all." Teachers may be able to use a particular program in the classroom but may find that it suits some students better than others. At all grade levels, but particularly in kindergarten and the early grades, children are known to vary greatly in the skills they bring to school. Some children will already know letter-sound correspondences, and some will even be able to decode words, while others will have little or no letter knowledge. Teachers should be able to assess the needs of the individual students and tailor instruction to meet specific needs. However, it is more common for phonics programs to present a fixed sequence of lessons scheduled from the beginning to the end of the school year. In light of this, teachers need to be flexible in their phonics instruction in order to adapt it to individual student needs.

Children who have already developed phonics skills and can apply them appropriately in the reading process do not require the same level and intensity of phonics instruction provided to children at the initial phases of reading acquisition. Thus, it will also be critical to determine objectively the ways in which systematic phonics instruction can be optimally incorporated and integrated in complete and balanced programs of reading instruction. Part of this effort should be directed at preservice and inservice education to provide teachers with decisionmaking frameworks to guide their selection, integration, and implementation of phonics instruction within a complete reading program.

Teachers must understand that systematic phonics instruction is only one component—albeit a necessary component—of a total reading program; systematic phonics instruction should be integrated with other reading instruction in phonemic awareness, fluency, and comprehension strategies to create a complete reading program. While most teachers and educational decisionmakers recognize this, there may be a tendency in some classrooms, particularly in 1st grade, to allow phonics to become the dominant component, not only in the time devoted to it, but also in the significance attached. It is important not to judge children's reading competence solely on the basis of their phonics skills and not to devalue their interest in books because they cannot decode with complete accuracy. It is also critical for teachers to understand that systematic phonics instruction can be provided in an entertaining, vibrant, and creative manner.

Systematic phonics instruction is designed to increase accuracy in decoding and word recognition skills, which in turn facilitate comprehension. However, it is again important to note that fluent and automatic application of phonics skills to text is another critical skill that must be taught and learned to maximize oral reading and reading comprehension. This issue again underscores the need for teachers to understand that while phonics skills are necessary in order to learn to read, they are not sufficient in their own right. Phonics skills must be integrated with the development of phonemic awareness, fluency, and text reading comprehension skills.

Fluency

Fluent readers are able to read orally with speed, accuracy, and proper expression. Fluency is one of several critical factors necessary for reading comprehension. Despite its importance as a component of skilled reading, fluency is often neglected in the classroom. This is unfortunate. If text is read in a laborious and inefficient manner, it will be difficult for the child to remember what has been read and to relate the ideas expressed in the text to his or her background knowledge. Recent research on the efficacy of certain approaches to teaching fluency has led to increased recognition of its importance in the classroom and to changes in instructional practices.

Reading practice is generally recognized as an important contributor to fluency. Two instructional approaches, each of which has several variations, have typically been used to teach reading fluency. One, guided repeated oral reading, encourages students to read passages orally with systematic and explicit guidance and feedback from the teacher. The other, independent silent reading, encourages students to read silently on their own, inside and outside the classroom, with minimal guidance or feedback.

Guided Oral Reading

The NRP conducted an initial series of electronic literature searches and identified 364 studies potentially relevant to the effects of guided oral reading instructional practices. Of these, 16 studies met the NRP research methodology criteria and were included in a meta-analysis, and 21 additional studies met the criteria but could not be included in the meta-analysis—although they were used in the qualitative interpretation of the efficacy of these instructional methods.

Findings and Determinations

On the basis of a detailed analysis of the available research that met NRP methodological criteria, the Panel concluded that guided repeated oral reading procedures that included guidance from teachers, peers, or parents had a significant and positive impact on word recognition, fluency, and comprehension across a range of grade levels. These studies were conducted in a variety of classrooms in both regular and special education settings with teachers using widely available instructional materials. This suggests the classroom readiness of guided oral reading and repeated reading procedures. These results also apply to all students—good readers as well as those experiencing reading difficulties. Nevertheless, there were important gaps in the research. In particular, the Panel could find no multiyear studies providing information on the relationship between guided oral reading and the emergence of fluency.

Independent Silent Reading

There has been widespread agreement in the literature that encouraging students to engage in wide, independent, silent reading increases reading achievement. Literally hundreds of correlational studies find that the best readers read the most and that poor readers read the least. These correlational studies suggest that the more that children read, the better their fluency, vocabulary, and comprehension. However, these findings are correlational in nature, and correlation does not imply causation. No doubt, it could be that the more that children read, the more their reading skills improve, but it is also possible that better readers simply choose to read more.

In order to address this issue of causation, the panel examined the specific impact that encouraging students to read more has on fluency, vocabulary development, and reading comprehension. The studies that were identified that address this issue were characterized by three major features. First, the studies emphasized silent reading procedures with students reading on their own with little or no specific feedback. Second, the studies did not directly assess fluency or the actual increase in the amount of reading due to the instructional procedures. Rather, only changes in vocabulary and/or comprehension were typically measured as outcomes rather than increases in fluency that could be expected from the increased reading practice. Third, very few studies that examined the effect of independent silent reading on reading achievement could meet the NRP research review methodology criteria (n = 14), and these studies varied widely in their methodological quality and the reading outcome variables measured. Thus, a meta-analysis could not be conducted. Rather, the 14 studies were examined individually and in detail to identify converging trends and findings in the data.

Findings and Determinations

With regard to the efficacy of having students engage in independent silent reading with minimal guidance or feedback, the Panel was unable to find a positive relationship between programs and instruction that

encourage large amounts of independent reading and improvements in reading achievement, including fluency. In other words, even though encouraging students to read more is intuitively appealing, there is still not sufficient research evidence obtained from studies of high methodological quality to support the idea that such efforts reliably increase how much students read or that such programs result in improved reading skills. Given the extensive use of these techniques, it is important that such research be conducted.

It should be made clear that these findings do not negate the positive influence that independent silent reading *may* have on reading fluency, nor do the findings negate the possibility that wide independent reading significantly influences vocabulary development and reading comprehension. Rather, there are simply not sufficient data from well-designed studies capable of testing questions of causation to substantiate causal claims. The available data do suggest that independent silent reading is not an effective practice when used as the *only* type of reading instruction to develop fluency and other reading skills, particularly with students who have not yet developed critical alphabetic and word reading skills. In sum, methodologically rigorous research designed to assess the specific influences that independent silent reading practices have on reading fluency and other reading skills and the motivation to read has not yet been conducted.

Comprehension

Comprehension is critically important to the development of children's reading skills and therefore to the ability to obtain an education. Indeed, reading comprehension has come to be the "essence of reading" (Durkin, 1993), essential not only to academic learning in all subject areas but to lifelong learning as well. In carrying out its analysis of the extant research in reading comprehension, the NRP noted three predominant themes in the research on the development of reading comprehension skills. First, reading comprehension is a complex cognitive process that cannot be understood without a clear description of the role that vocabulary development and vocabulary instruction play in the understanding of what has been read. Second, comprehension is an active process that requires an intentional and thoughtful interaction between the reader and the text. Third, the preparation of teachers to better equip students to develop and apply reading comprehension strategies to enhance understanding is intimately linked to students' achievement in this area. Because these three themes serve as the foundation for understanding how best to help teachers develop students' comprehension abilities, the extant research relevant to vocabulary instruction, to text comprehension instruction, and to the preparation of teachers to teach reading comprehension strategies was examined in detail by the NRP. The major findings and determinations of the Panel for each of these three subareas are provided next.

Vocabulary Instruction

The importance of vocabulary knowledge has long been recognized in the development of reading skills. As early as 1924, researchers noted that growth in reading power means continuous growth in word knowledge (Whipple, 1925). Vocabulary is critically important in oral reading instruction. There are two types of vocabulary—oral and print. A reader who encounters a strange word in print can decode the word to speech. If it is in the reader's oral vocabulary, the reader will be able to understand it. If the word is not in the reader's oral vocabulary, the reader will have to determine the meaning by other means, if possible. Consequently, the larger the reader's vocabulary (either oral or print), the easier it is to make sense of the text.

To determine how vocabulary can best be taught and related to the reading comprehension process, the NRP examined more than 20,000 research citations identified through electronic and manual literature searches. From this set, citations were removed if they did not meet prespecified criteria: if they were not reports of research, if they were not reporting experimental or quasi-experimental studies, if they

were not published in English, or if they dealt exclusively with learning disabled or other special populations, including second-language learners. Comprehensive review of the remaining set of studies according to the NRP review criteria identified 50 studies for further evaluation. Further analysis and coding of these studies indicated that a formal meta-analysis could not be conducted because there was a small number of research studies in vocabulary instruction dealing with a relatively large number of variables. There are recent published meta-analyses for some selected variables, and it was decided not to duplicate those efforts. Also, a substantial amount of published research on vocabulary instruction did not meet NRP research methodology criteria. Because the Panel wanted to glean as much information as possible from the studies identified in the searches, the vocabulary instruction database was reviewed for trends across studies, even though formal meta-analyses could not be conducted. Fifty studies dating from 1979 to the present were reviewed in detail. There were 21 different methods represented in these studies.

Findings and Determinations

The studies reviewed suggest that vocabulary instruction does lead to gains in comprehension, but that methods must be appropriate to the age and ability of the reader. The use of computers in vocabulary instruction was found to be more effective than some traditional methods in a few studies. It is clearly emerging as a potentially valuable aid to classroom teachers in the area of vocabulary instruction. Vocabulary also can be learned incidentally in the context of storybook reading or in listening to others. Learning words before reading a text also is helpful. Techniques such as task restructuring and repeated exposure (including having the student encounter words in various contexts) appear to enhance vocabulary development. In addition, substituting easy words for more difficult words can assist low-achieving students.

The findings on vocabulary yielded several specific implications for teaching reading. First, vocabulary should be taught both directly and indirectly. Repetition and multiple exposures to vocabulary items are important. Learning in rich contexts, incidental learning, and use of computer technology all enhance the acquisition of vocabulary. Direct instruction should include task restructuring as necessary and should actively engage the student. Finally, dependence on a single vocabulary instruction method will not result in optimal learning.

While much is known about the importance of vocabulary to success in reading, there is little research on the best methods or combinations of methods of vocabulary instruction and the measurement of vocabulary growth and its relation to instruction methods.

Text Comprehension Instruction

Comprehension is defined as "intentional thinking during which meaning is constructed through interactions between text and reader" (Harris & Hodges, 1995). Thus, readers derive meaning from text when they engage in intentional, problem solving thinking processes. The data suggest that text comprehension is enhanced when readers actively relate the ideas represented in print to their own knowledge and experiences and construct mental representations in memory.

The rationale for the explicit teaching of comprehension skills is that comprehension can be improved by teaching students to use specific cognitive strategies or to reason strategically when they encounter barriers to understanding what they are reading. Readers acquire these strategies informally to some extent, but explicit or formal instruction in the application of comprehension strategies has been shown to be highly effective in enhancing understanding. The teacher generally demonstrates such strategies for students until the students are able to carry them out independently.

The literature search identified 453 studies that addressed issues and topics relevant to text comprehension since 1980. Studies published between 1970 and 1979 were added if they were of particular relevance, resulting in 481 studies that were initially reviewed. Of these, 205 studies met the general NRP methodological criteria and were then classified into instructional categories based on the kind of instruction used. Application of the more specific review criteria precluded formal meta-analyses because of the large variation in methodologies and implementations used. The Panel found few research studies that met all NRP research methodology criteria. Nevertheless, the Panel employed the NRP criteria to the maximum extent possible in its examination of this body of literature. (See the Comprehension section of the Report of the National Reading Panel: Reports of the Subgroups.)

In its review, the Panel identified 16 categories of text comprehension instruction of which 7 appear to have a solid scientific basis for concluding that these types of instruction improve comprehension in non-impaired readers. Some of these types of instruction are helpful when used alone, but many are more effective when used as part of a multiple-strategy method. The types of instruction are:

- Comprehension monitoring, where readers learn how to be aware of their understanding of the material;
- Cooperative learning, where students learn reading strategies together;
- Use of graphic and semantic organizers (including story maps), where readers make graphic representations of the material to assist comprehension;
- Question answering, where readers answer questions posed by the teacher and receive immediate feedback;
- Question generation, where readers ask themselves questions about various aspects of the story;
- Story structure, where students are taught to use the structure of the story as a means of helping them recall story content in order to answer questions about what they have read; and
- Summarization, where readers are taught to integrate ideas and generalize from the text information.

Findings and Determinations

In general, the evidence suggests that teaching a combination of reading comprehension techniques is the most effective. When students use them appropriately, they assist in recall, question answering, question generation, and summarization of texts. When used in combination, these techniques can improve results in standardized comprehension tests.

Nevertheless, some questions remain unanswered. More information is needed on ways to teach teachers how to use such proven comprehension strategies. The literature also suggests that teaching comprehension in the context of specific academic areas—for example, social studies—can be effective. If this is true of other subject areas, then it might be efficient to teach comprehension as a skill in content areas.

Questions remain as to which strategies are most effective for which age groups. More research is necessary to determine whether the techniques apply to all types of text genres, including narrative and expository texts, and whether the level of difficulty of the texts has an impact on the effectiveness of the strategies. Finally, it is critically important to know what teacher characteristics influence successful instruction of reading comprehension.

Teacher Preparation and Comprehension Strategies Instruction

Teaching reading comprehension strategies to students at all grade levels is complex. Teachers not only must have a firm grasp of the content presented in text, but also must have substantial knowledge of the strategies

themselves, of which strategies are most effective for different students and types of content and of how best to teach and model strategy use.

Research on comprehension strategies has evolved dramatically over the last 2 decades. Initially, investigators focused on teaching one strategy at a time; later studies examined the effectiveness of teaching several strategies in combination. However, implementation of this promising approach has been problematic. Teachers must be skillful in their instruction and be able to respond flexibly and opportunistically to students' needs for instructive feedback as they read.

The initial NRP search for studies relevant to the preparation of teachers for comprehension strategy instruction provided 635 citations. Of these, only four studies met the NRP research methodology criteria. Hence, the number of studies eligible for further analysis precluded meta-analysis of the data derived from these investigations. However, because there were only four studies, the NRP was able to review them in detail. The studies investigate two major approaches: Direct Explanation and Transactional Strategy Instruction.

The Direct Explanation approach focuses on the teacher's ability to explain explicitly the reasoning and mental processes involved in successful reading comprehension. Rather than teach specific strategies, teachers help students (1) to view reading as a problem solving task that necessitates the use of strategic thinking, and (2) to learn to think strategically about solving comprehension problems. For example, teachers are taught that they could teach students the skill of finding the main idea by casting it as a problemsolving task and reasoning about it strategically.

Transactional Strategy Instruction also emphasizes the teacher's ability to provide explicit explanations of thinking processes. Further, it emphasizes the ability of teachers to facilitate student discussions in which students collaborate to form joint interpretations of text and acquire a deeper understanding of the mental and cognitive processes involved in comprehension.

Findings and Determinations

The four studies (two studies for each approach) demonstrated that teachers could be instructed in these methods. Teachers required instruction in explaining what they are teaching, modeling their thinking processes, encouraging student inquiry, and keeping students engaged. Data from all four studies indicated clearly that in order for teachers to use strategies effectively, extensive formal instruction in reading comprehension is necessary, preferably beginning as early as preservice.

More research is needed to address the following questions. Which components of teacher preparation are most effective? Can reading comprehension strategies be successfully incorporated into content area instruction? How can the effectiveness of strategies be measured in an optimal manner? Can strategies be taught as early as grades 1 and 2, when children also are trying to master phonics, word recognition, and fluency? How can teachers be taught to provide the most optimal instruction?

Teacher Education and Reading Instruction

Recent developments such as class size reduction and the writing of standards suggest the growing importance of teacher education on learning outcomes. In addition, the National Reading Panel decided to focus on this area because during its regional meetings speakers expressed intense interest in the quality and importance of teacher education.

In teacher education programs, preservice teachers generally acquire knowledge through supervised teaching and through coursework in theory and methods. Continuing education for practicing teachers comes from professional development, also called inservice education. The NRP analysis on this topic was guided by three primary questions: How are teachers taught to teach reading? What does research

show about the effectiveness of this instruction? How can research be applied to improve teacher development? The initial literature search by the Panel identified more than 300 articles. A total of 32 studies met the methodological NRP criteria: 11 preservice and 21 inservice. No meta-analysis was conducted because the range of variables and theoretical positions was too large for the limited number of studies.

Findings and Determinations

As indicated by the NRP's examination of the literature, only a small number of experimental studies have been published about the effectiveness of preservice and inservice teacher education. For conclusions to be drawn about the effectiveness of teacher education, information on both teacher and student outcomes must be reported. Preservice research, however, only measured teacher outcomes, whereas ideally both short- and long-term teacher and student outcomes should be observed. With respect to research on inservice education, only about one-half measured student outcomes as well as teacher outcomes.

Generally the results indicated that inservice professional development produced significantly higher student achievement. There were few studies of the long-term maintenance of the gains. While there were only a small number of studies, almost all of them showed positive effects on teaching. However, there were too few studies on specific variables to allow the Panel to draw specific conclusions about the content of preservice education.

More information is needed in several areas. What is the optimal combination of preservice and inservice education, and what are the effects of preservice experience on inservice performance? What is the appropriate length of inservice and preservice education? What are the best ways to assess the effectiveness of teacher education and professional development? How can teachers optimally be supported over the long term to ensure sustained implementation of new methods and to ensure student achievement? The relationship between the development of standards and teacher education is also an important gap in current knowledge.

Computer Technology and Reading Instruction

Until recently, computers were not considered capable of delivering reading instruction effectively. They could not comprehend oral reading and judge its accuracy. They also were unable to accept free-form responses to comprehension questions, so their use had to rely primarily on multiple-choice formats. Today, the situation is much improved. New computers have speech recognition capabilities as well as many multimedia presentation functions. Developments in the Internet, with possibilities of linking schools and instruction, have further increased interest in technology as a teaching device. Computer technology is different from other areas the NRP analyzed. It cannot be studied independently of instructional content and is not an instructional method in itself. Thus, computer technology must be examined for its ability to deliver instruction, for example, in vocabulary or in phonemic awareness.

Because this is a relatively new field, the number of studies published in this area is small. Only 21 studies met the NRP research methodology criteria.

Findings and Determinations

Although it is difficult to draw conclusions from these studies, it is possible to make some general statements. First, all the studies report positive results, suggesting that it is possible to use computer technology for reading instruction. The seven studies that reviewed the addition of speech to computer-presented text indicate that this may be a promising use of technology in reading instruction.

Two other trends show promise. The use of hypertext (highlighted text that links to underlying definitions or supporting or related text, almost like an electronic footnote), while technically not reading instruction, may have an instructional advantage. Second, the use of computers as word processors may be very useful, given that reading instruction is most effective when combined with writing instruction.

Striking in its absence is research on the incorporation of Internet applications to reading instruction.

Research also is needed on the value of speech recognition as a technology and the use of multimedia presentations in reading instruction.

In sum, the Panel is encouraged by the reported successes in the use of computer technology for reading instruction, but relatively few specific instructional applications can be gleaned from the research. Many questions still need to be addressed.

Next Steps

As part of its Congressional charge, the NRP was directed to assess the effectiveness of various approaches to teaching children to read and to further indicate the extent to which effective approaches were ready for application in classroom settings. The instructional topics of alphabetics (phonemic awareness and phonics), fluency, comprehension (vocabulary instruction and text comprehension), teacher education and reading instruction, and computer technology and reading instruction addressed in this Report were selected by the Panel from a candidate list of 35 topics generated from Panel members' own expertise, from the report of the National Research Council on Preventing Reading Difficulties in Young Children (Snow, Burns, & Griffin, 1998), and from the input the Panel received in its regional hearings. Several additional factors contributed to the consensus decision to limit the number of topics that could be addressed and to evaluate the research literature relevant to these specific topics. These factors included (1) the hypothesized role that these topics play in reading instruction; (2) the availability of well-designed experimental or quasi-experimental studies of instructional effectiveness for each of these topic areas versus other topic areas; (3) the immensity of the research literature in reading development and reading instruction; and (4) constraints on time and Panel resources.

The Panel regrets that it could not evaluate all of the reading instructional topics that were identified by Panel members as well as by parents, educators, and policymakers at the regional meetings. The Panel emphasizes that omissions of topics such as the effects of predictable and decodable text formats on beginning reading development, motivational factors in learning to read, and the effects of integrating reading and writing, to name a few, are not to be interpreted as determinations of unimportance or ineffectiveness. Indeed, each of the reports of the subgroups identifies areas for future research. These can serve as checklists of important research opportunities for further analyses and evaluations of the kind conducted by the Panel on this first set of topics.

It is the Panel's fervent hope that future evaluations of important reading research topics will include an analysis and assessment of correlational, descriptive, and qualitative studies that inform our understanding of the developmental reading process, and a determination of what instructional implications can be drawn from them. Moreover, it will be critical to understand better how quantitative, hypothesis-driven studies can best be integrated with qualitative approaches to obtain maximum reliability and ecological validity. Likewise, it will be critical to identify the most important methodological features inherent in qualitative and descriptive research approaches that lead to the collection of trustworthy evidence. Thus, the Panel recommends that the evaluation of these types of qualitative research approaches, methods, and evidence be guided by the development of a comparable methodologically rigorous review process similar to that employed by the NRP with procedures and criteria designated *a priori* and applied within an open and public forum.

With this information as background, it is clear to the Panel that at least four major tasks remain in developing a science of reading development and reading instruction. First, where possible, there should be meta-analyses of existing experimental or quasi-experimental research in topic areas not addressed by the NRP. Second, additional experimental research should be conducted on questions unanswered by the Panel's analyses of the topics it did cover. Third, there should be an exhaustive and objective analysis of correlational, descriptive, and qualitative studies relevant to reading development and reading instruction that is carried out with methodological rigor following pre-established criteria. Fourth, experimental research should be initiated to test those

hypotheses derived from existing correlational, descriptive, and qualitative research meeting high methodological standards.

Following are three illustrative examples of important reading research opportunities.

- Student Populations. An important question is whether students with learning disabilities have distinctive instructional needs and whether they benefit from instructional techniques that are different from those that are optimal for other low-achieving (non-disabled) students. The Panel was able to address this question with respect to phonemic awareness and phonics instructional programs and techniques. It found that both types of students benefit from similar phonemic awareness and phonics instructional programs and techniques. Because of the limited amount of research available, the Panel could not answer this question with respect to instructional programs and techniques aimed at developing reading fluency and comprehension. These important comparisons should be the focus of future research.
- Teacher Education. The primary purpose of teacher education research is to inform the effective practice of classroom teachers in order to improve student performance. Rigorous experimental and qualitative research that defines and characterizes effective teaching methodologies that demonstrate improved student performance is limited. This persistent and major gap in the extant knowledge base must be addressed. Efforts should be made to answer the important questions in this critical area.
- Uses of Technology in Teaching Reading. Here again, credible experimental and qualitative research is lacking. This is understandable in light of the recent development of the relevant technology and its application to reading instruction and student learning. Nevertheless, the Panel believes that this is an important and essentially unexplored field.

Reflections

The findings and determinations of the NRP reflect a focused and persistent effort on the part of the Panel to contribute reliable, valid, and trustworthy information to the body of knowledge that is leading to a better scientific understanding of reading development and reading instruction. In carrying out its Congressional charge, the Panel was able to first develop, and then to apply a methodologically rigorous research review process and protocol and to do so within an open and public forum. The *a priori* establishment of research review criteria, the systematic evaluation process, and the openness to public scrutiny at all times ensured that the evidence ultimately evaluated by the Panel met well-established objective scientific standards. This process also serves as a model for future evaluations of evidence obtained experimentally on other topics relevant to reading as well as for studies employing nonexperimental methodologies.

The work of the NRP builds on existing knowledge about what types of skills children need to acquire to become independent readers. Specifically, the Panel addresses the evidence about what those skills are and adds further knowledge about how those skills are best taught to beginning readers who vary in initial reading-related abilities. The Panel identified a number of instructional approaches, methods, and strategies that hold substantial promise for application in the classroom at this time. Specifically, the Report of the National Reading Panel: Reports of the Subgroups includes specific findings that can be useful in helping teachers develop instructional applications with students. Moreover, the Reports of the Subgroups provides extensive references that teachers can locate for instructional ideas and guidance. In addition, the Panel identified areas where significantly greater research effort is needed, and where the quality of the research efforts must improve in order to determine objectively the effectiveness of different types of reading instruction. Significantly, the Panel has reached a series of positive conclusions about several areas of instructional research through a rigorous and open process. We are confident that the determinations made by the Panel in this regard will benefit children, teachers, and educational policymakers.

References

Cooper, H., & Hedges, L.V. (1994). The handbook of research synthesis. New York: Russell Sage Foundation.

Durkin, D. (1993). Teaching them to read (6th ed.). Boston, MA: Allyn & Bacon.

Harris, T., & Hodges, R. (Eds.). (1995). The literacy dictionary (p. 207). Newark, DE: International Reading Association.

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

Whipple, G. (Ed.). (1925). The Twenty-fourth Yearbook of the National Society for the Study of Education: Report of the National Committee on Reading. Bloomington, IL: Public School Publishing Company.

ADDENDUM

Methodology: Processes Applied to the Selection, Review, and Analysis of Research Relevant to Reading Instruction

In an important action critical to its Congressional charge, the NRP elected to develop and adopt a set of rigorous research methodological standards. These standards, which are defined in this section, guided the screening of the research literature relevant to each topic area addressed by the Panel. This screening process identified a final set of experimental or quasi-experimental research studies that were then subjected to detailed analysis. The evidence-based methodological standards adopted by the Panel are essentially those normally used in research studies of the efficacy of interventions in psychological and medical research. These include behaviorally based interventions, medications, or medical procedures proposed for use in the fostering of robust health and psychological development and the prevention or treatment of disease. It is the view of the Panel that the efficacy of materials and methodologies used in the teaching of reading and in the prevention or treatment of reading disabilities should be tested no less rigorously. However, such standards have not been universally accepted or used in reading education research. Unfortunately, only a small fraction of the total reading research literature met the Panel's standards for use in the topic analyses.

With this as background, the Panel understood that criteria had to be developed as it considered which research studies would be eligible for assessment. There were two reasons for determining such guidelines or rules *a priori*. First, the use of common search, selection, analysis, and reporting procedures would ensure that the Panel's efforts could proceed, not as a diverse collection of independent—and possibly uneven—synthesis papers, but as parts of a greater whole. The use of common procedures permitted a more unified presentation of the combined methods and findings. Second, the amount of research synthesis that had to be accomplished was substantial. Consequently, the Panel had to work in diverse subgroups to identify, screen, and evaluate the relevant research to complete their respective reports. Moreover, the Panel also had to arrive at findings that all or nearly all of the members of the NRP could endorse. Common procedures, grounded in scientific principles, helped the Panel to reach final agreements.

Search Procedures

Each subgroup conducted a search of the literature using common procedures, describing in detail the basis and rationale for its topical term selections, the strategies employed for combining terms or delimiting searches, and the search procedures used for each topical area.

Each subgroup limited the period of time covered by its searches on the basis of relative recentness and how much literature the search generated. For example, in some cases it was decided to limit the years searched to the number of most recent years that would identify between 300 and 400 potential sources. This scope could be expanded in later iterations if it appeared that the nature of the research had changed qualitatively over time, if the proportion of useable research identified was small (e.g., less than 25%), or if the search simply represented too limited a proportion of the total set of identifiable studies. Although the number of years searched varied among subgroup topics, decisions regarding the number of years to be searched were made in accord with shared criteria.

The initial criteria were established to focus the efforts of the Panel. First, any study selected had to focus directly on children's reading development from preschool through grade 12. Second, the study had to be published in English in a refereed journal. At a minimum, each subgroup searched both PsycINFO and ERIC databases for studies meeting these initial criteria. Subgroups could, and did, use additional

databases when appropriate. Although the use of a minimum of two databases identified duplicate literature, it also afforded the opportunity to expand perspective and locate articles that would not be identifiable through a single database.

Identification of each study selected was documented for the record, and each was assigned to one or more members of the subgroup, who examined the title and abstract. Based on this examination, the subgroup member(s) determined, if possible at this stage, whether the study addressed issues within the purview of the research questions being investigated. If it did not, the study was excluded and the reason(s) for the exclusion were detailed and documented for the record. If it did address reading instructional issues relevant to the Panel's selected topic areas, the study underwent further examination.

Following initial examination, if the study had not been excluded in accord with the preceding criteria, the full study report was located and examined in detail to determine whether the following criteria were met:

- Study participants must be carefully described (age, demographic, cognitive, academic, and behavioral characteristics);
- Study interventions must be described in sufficient detail to allow for replicability, including how long the interventions lasted and how long the effects lasted;
- Study methods must allow judgments about how instruction fidelity was insured; and
- Studies must include a full description of outcome measures.

These criteria for evaluating research literature are widely accepted by scientists in disciplines involved in medical, behavioral, and social research. The application of these criteria increases the probability that objective, rigorous standards were used and that therefore the information obtained from the studies would contribute to the validity of any conclusions drawn.

If a study did not meet these criteria or could not be located, it was excluded from subgroup analysis and the reason(s) for its exclusion detailed and documented for the record. If the study was located and met the criteria, the study became one of the subgroup's core working set of studies. The core working sets of studies gathered by the subgroups were then coded as described below and then analyzed to address the questions posed in the introduction and in the charge to the Panel.

If a core set of studies identified by the subgroup was insufficient to answer critical instructional questions, less recent studies were screened for eligibility for, and inclusion in, the core working sets of studies. This second search used the reference lists of all core studies and known literature reviews. This process identified cited studies that could meet the Panel's methodological criteria for inclusion in the subgroups' core working sets of studies. Any second search was described in detail and applied precisely the same search, selection, exclusion, and inclusion criteria and documentation requirements as were applied in the subgroups' initial searches.

Manual searches, again applying precisely the same search, selection, exclusion, and inclusion criteria and documentation requirements as were applied in the subgroups' electronic searches, were also conducted to supplement the electronic database searches. Manual searching of recent journals that publish research on specific NRP subgroup topics was performed to compensate for the delay in appearance of these journal articles in the electronic databases. Other manual searching was carried out in relevant journals to include eligible articles that should have been selected, but were missed in electronic searches.

Source of Publications: The Issue of Refereed and Non-Refereed Articles

The subgroup searches focused exclusively on research that had been published or had been scheduled for publication in refereed (peer-reviewed) journals. The Panel reached consensus that

determinations and findings for claims and assumptions guiding instructional practice depended on such studies. Any search or review of studies that had not been published through the peer review process but was consulted in any subgroup's review was treated as separate and distinct from evidence drawn from peer reviewed sources (i.e., in an appendix) and is not referenced in the Panel's report. These non-peer-reviewed data were treated as preliminary/pilot data that might illuminate potential trends and areas for future research. Information derived in whole or in part from such studies was not to be represented at the same level of certainty as findings derived from the analysis of refereed articles.

Types of Research Evidence and Breadth of Research Methods Considered

Different types of research (e.g., descriptive-interpretive, correlational, experimental) lay claim to particular warrants, and these warrants differ markedly. The Panel felt that it was important to use a wide range of research, but that the research be used in accordance with the purposes and limitations of the various research types.

To make a determination that any instructional practice could be or should be adopted widely to improve reading achievement requires that the belief, assumption, or claim supporting the practice is causally linked to a particular outcome. The highest standard of evidence for such a claim is the experimental study, in which it is shown that treatment can make such changes and effect such outcomes. Sometimes when it is not feasible to do a randomized experiment, a quasi-experimental study is conducted. This type of study provides a standard of evidence that, while not as high, is acceptable, depending on the study design.

To sustain a claim of effectiveness, the Panel felt it necessary that there be experimental or quasi-experimental studies of sufficient size or number, and scope (in terms of population served), and that these studies be of moderate to high quality. When there were too few studies of this type or they were too narrowly cast or they were of marginally acceptable quality, then it was essential that the Panel have substantial correlational or descriptive studies that concurred with the findings if a claim was to be sustained. No claim could be determined on the basis of descriptive or correlational research alone. The use of these procedures increased the possibility of reporting findings with a high degree of internal validity.

Coding of Data

Characteristics and outcomes of each study that met the screening criteria described above were coded and analyzed, unless otherwise authorized by the Panel. The data gathered in these coding forms were the information submitted to the final analyses. The coding was carried out in a systematic and reliable manner.

The various subgroups relied on a common coding form developed by a working group of the Panel's scientist members and modified and endorsed by the Panel. However, some changes could be made to the common form by the various subgroups for addressing different research issues. As coding forms were developed, any changes to the common coding form were shared with and approved by the Panel to ensure consistency across various subgroups.

Unless specifically identified and substantiated as unnecessary or inappropriate by a subgroup and agreed to by the Panel, each form for analyzing studies was coded for the following categories:

1. Reference

- Citation (standard APA format)
- How this paper was found (e.g., search of named database, listed as reference in another empirical paper or review paper, manual search of recent issues of journals)
- Narrative summary that includes distinguishing features of this study

2. Research Question: The general umbrella question that this study addresses

3. Sample of Student Participants

- States or countries represented in sample
- Number of different schools represented in sample
- Number of different classrooms represented in sample
- Number of participants (total, per group)
- Age
- Grade
- Reading levels of participants (prereading, beginning, intermediate, advanced)
- Whether participants were drawn from urban, suburban, or rural settings
- List any pretests that were administered prior to treatment
- List any special characteristics of participants including the following if relevant:
- Socioeconomic status (SES)
- Ethnicity
- Exceptional learning characteristics, such as:
  - Learning disabled
  - Reading disabled
  - Hearing impaired
- English language learners (ELL); also known as limited English proficient (LEP) students
- Explain any selection restrictions that were applied to limit the sample of participants (e.g., only those low in phonemic awareness were included)
- Contextual information: concurrent reading instruction that participants received in their classrooms during the study
  - Was the classroom curriculum described in the study? (code = yes/no)
  - Describe the curriculum
- Describe how sample was obtained:
  - Schools or classrooms or students were selected from the population of those available
  - Convenience or purposive sample
  - Not reported
  - Sample was obtained from another study (specify study)
- Attrition:
  - Number of participants lost per group during the study
  - Was attrition greater for some groups than for others? (yes/no)

4. Setting of the Study

- Classroom
- Laboratory
- Clinic
- Pullout program (e.g., Reading Recovery©)
- Tutorial

5. Design of Study

- Random assignment of participants to treatments (randomized experiment)
  - With vs. without a pretest
- Nonequivalent control group design (quasi-experiment), e.g., existing groups assigned to treatment or control conditions, no random assignment
  - With vs. without matching or statistical control to address nonequivalence issue
- One-group repeated measure design (i.e., one group receives multiple treatments, considered a quasi-experiment)
  - Treatment components administered in a fixed order vs. order counterbalanced across subgroups of participants
- Multiple baseline (quasi-experiment)
  - Single-subject design
  - Aggregated-subjects design

6. Independent Variables

a. Treatment Variables

- Describe all treatments and control conditions; be sure to describe nature and components of reading instruction provided to control group.
- For each treatment, indicate whether instruction was explicitly or implicitly delivered and, if explicit instruction, specify the unit of analysis (sound-symbol; onset/rime; whole word) or specific responses taught. [Note: If this category is omitted in the coding of data, justification must be provided.]
- If text is involved in treatments, indicate difficulty level and nature of texts used
- Duration of treatments (given to students)
  - Minutes per session
  - Sessions per week
  - Number of weeks
- Was trainers' fidelity in delivering treatment checked? (yes/no)
- Properties of teachers/trainers
- Number of trainers who administered treatments
- Teacher/student ratio: Number of trainers to number of participants
- Type of trainer (classroom teacher, student teacher, researcher, clinician, special education teacher, parent, peer, other)
- List any special qualifications of trainers
- Length of training given to trainers
- Source of training
- Assignment of trainers to groups:
  - Random
  - Choice/preference of trainer
  - All trainers taught all conditions
- Cost factors: List any features of the training such as special materials or staff development or outside consultants that represent potential costs

b. Moderator Variables

- List and describe other nontreatment independent variables included in the analyses of effects (e.g., attributes of participants, properties or types of text)

7. Dependent (Outcome) Variables

- List processes that were taught during training and measured during and at the end of training
- List names of reading outcomes measured
  - Code each as standardized or investigator-constructed measure
  - Code each as quantitative or qualitative measure
  - For each, is there any reason to suspect low reliability? (yes/no)
- List time points when dependent measures were assessed

8. Nonequivalence of groups

- Any reason to believe that treatment/control group might not have been equivalent prior to treatments? (yes/no)
- Were steps taken in statistical analyses to adjust for any lack of equivalence? (yes/no)

9. Result (for each measure)

- Record the name of the measure
- Record whether the difference—treatment mean minus control mean—is positive or negative
- Record the value of the effect size including its sign (+ or -)
- Record the type of summary statistics from which the effect size was derived
- Record number of people providing the effect size information

10. Coding Information

- Record length of time to code study
- Record name of coder

If text was a variable, the coding indicated what is known about the difficulty level and nature of the texts being used. Any use of special personnel to deliver an intervention, use of special materials, staff development, or other features of the intervention that represent potential cost were noted. Finally, various threats to reliability and internal or external validity

(group assignment, teacher assignment, fidelity of treatment, and confounding variables including equivalency of subjects prior to treatment and differential attrition) were coded. Each subgroup also coded additional items deemed appropriate or valuable to the specific question being studied by the subgroup members.

A study could be excluded at the coding stage only if it was found to have so serious a fundamental flaw that its use would be misleading. The reason(s) for exclusion of any such study was detailed and documented for the record. When quasi-experimental studies were selected, it was essential that each study included both pre-treatment and post-treatment evaluations of performance and that there was a comparison group or condition.

Each subgroup conducted an independent re-analysis of a randomly designated 10% sample of studies. Absolute rating agreement was calculated for each category (not for forms). If absolute agreement fell below 0.90 for any category for occurrence or nonoccurrence agreement, the subgroup took some action to improve agreement (e.g., multiple readings with resolution, improvements in coding sheet).

Upon completion of the coding for recently published studies, a letter was sent to the first author of the study requesting any missing information. Any information that was provided by authors was added to the database.

After its search, screening, and coding, a subgroup determined whether for a particular question or issue a meaningful meta-analysis could be completed or whether it was more appropriate to conduct a literature analysis of that issue or question without meta-analysis, incorporating all of the information gained. The full Panel reviewed and approved or modified each decision.

Data Analysis

When appropriate and feasible, effect sizes were calculated for each intervention or condition in experimental and quasi-experimental studies. The subgroups used the standardized mean difference formula as the measure of treatment effect. The formula was:

$(M_t - M_c) / 0.5(sd_t + sd_c)$

where:

M_t is the mean of the treated group,

M_c is the mean of the control group,

sd_t is the standard deviation of the treated group, and

sd_c is the standard deviation of the control group.

When means and standard deviations were not available, the subgroups followed the guidelines for the calculation of effect sizes as specified by Cooper and Hedges (1994).

The subgroups weighted effect sizes by numbers of subjects in the study or comparison to prevent small studies from overwhelming the effects evident in large studies.

Each subgroup used median and/or average effect sizes when a study had multiple comparisons, and each subgroup only employed the comparisons that were specifically relevant to the questions under review by the subgroup.

Expected Outcomes

Analyses of effect sizes were undertaken with several goals in mind. First, overall effect sizes of related studies were calculated across subgroups to determine the best estimate of a treatment's impact on reading. These overall effects were examined with regard to their difference from zero (i.e., does the treatment have an effect on reading?), strength (i.e., if the treatment has an effect, how large is that effect?), and consistency (i.e., did the effect of the treatment vary

significantly from study to study?). Second, the Panel compared the magnitude of a treatment's effect under different methodological conditions, program contexts, program features, and outcome measures and for students with different characteristics. The appropriate moderators of a treatment's impact were drawn from the distinctions in studies recorded on the coding sheets. In each case, a statistical comparison was made to examine the impact of each moderator variable on average effect sizes for each relevant outcome variable. These analyses enabled the Panel to determine the conditions that alter a program's effects and the types of individuals for whom the program is most and least effective. Within-group average effect sizes were examined as were overall effect sizes for differences from zero and for strength. The analytic procedures were carried out using the techniques described by Cooper and Hedges (1994).

References

Cooper, H., & Hedges, L. V. (1994). The handbook of research synthesis. New York: Russell Sage Foundation.

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