

# CHAPTER 1

## Introduction and Methodology





---

## REPORTS OF THE SUBGROUPS

### 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.” The 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 was composed of 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 the Congress in February 1999. The information provided in the NRP Progress Report, the [Report of the National Reading Panel](#), and this [Report of the National Reading Panel: Reports of the Subgroups](#) reflects 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 subgroup 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 full report of each subgroup is included in this volume.) 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 repeated 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.



---

## REPORTS OF THE SUBGROUPS

### 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 to 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, however, 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; demographics; 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 ensured.
- 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 increased 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 was 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 subgroups 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 that 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 be 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 either 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, hand 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 setting
- 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) (Example: 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 participants to number of trainers
- 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 non-occurrence 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 in 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 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, 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 in 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.