

## ESSAY BOOK REVIEWS

*Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.*

By The National Reading Panel. 2000. Washington, DC: National Institute of Child Health and Human Development. NIH Publication No. 00-4769. Softcover.

*Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups.*

By The National Reading Panel. 2000. Washington, DC: National Institute of Child Health and Human Development. NIH Publication No. 00-4754. Softcover.

# The National Reading Panel Report

James W. Cunningham

*University of North Carolina at Chapel Hill, USA*

At the behest of the United States Congress in 1997, the Director of the National Institute of Child Health and Human Development (NICHD) and the U.S. Secretary of Education selected 14 persons to serve as a National Reading Panel (NRP). Most Panel members were reading researchers in various fields. All but two members held a doctorate. The Panel was charged to review and assess the research on teaching reading, with implications for both classroom practice and further research.

The report of the National Reading Panel was issued in two volumes. The first volume (00-4769) is a succinct summary of how the Panel came to be, the topics it chose to investigate, its procedures and methods, and its findings. The second volume (00-4754) contains the same introductory and methodological information, but presents at great length the work of each of the topical sub groups within the Panel. It is the second volume that one must read to fully understand the findings and recommendations for classroom practice and future research.

In this review, I refer to both volumes collectively as the NRP Report. Citations of the first volume contain only page numbers (e.g., p. 4); citations of the second volume contain a section number followed by page numbers because the second volume's pagination starts with 1 in each section (e.g., p. 3-13 means section 3, page 13 of the second volume). Some statements appear verbatim in both volumes.

### *The NRP's philosophy of science*

The NRP Report should be seen as a manifesto for a particular philosophy of science as much as a summary of particular research findings. Marks of the manifesto are not subtle and, indeed, begin on the cover. The subtitle of both volumes of the report asserts that the Panel has provided us with "an evidence-based assessment of the *scientific* research literature" (covers, emphasis added). The Methodological Overview of the first volume begins with the sentence, "In what may be its most important action, the Panel then developed and adopted a set of rigorous research methodological standards" (p. 5, emphasis added). In their Reflections, the Panel claims that its goal had been to contribute "to a better *scientific* understanding of reading development and reading instruction" (p. 21, emphasis added). Upon looking back at its completed work, it assures us that "the evidence ultimately evaluated by the Panel met well-established objective scientific standards" (p. 21, emphasis added).

The Report makes it clear that the methodological standards adopted by the Panel did not arise from the research literature on reading, but rather were imposed upon it. Panel members tell us that they developed their

criteria “a priori,” (p. 27; p. 1-5) and that “Unfortunately, only a small fraction of the total reading research literature met the Panel’s standards for use in the topic analyses” (p. 27; p. 1-5).

What are we to make of a report that so boldly lays claim to what science, rigor, and objectivity are in reading research, and first denigrates, then ignores, the preponderance of research literature in our field? Even though the NRP’s philosophy of science is implied, its consequences are not discussed, so making it explicit and discussing it here is important. The Panel members’ position about what kind of research is scientific fits within a historical philosophical context. To the extent that their views on science may affect how funding agencies, reviewers for journals and conference programs, and researchers conduct themselves, they have implications for the nature of future research in reading. If used to inform policy, their views on science will affect classroom reading experiences every day.

## Demarcation

The Panel members’ repeated and unapologetic appropriation of the term *science* to describe the results of their work places how they characterize their work in the subdomain of philosophy of science concerned with the demarcation problem. Positivism (Comte, 1830/1988) was an attempt to define science as knowledge with no vestige of theology or abstraction. Science was to be differentiated, or demarcated, from nonscience by being limited to beliefs that are so empirically supported they are certain or positive.

In the century after Comte’s first work, scientific practice demonstrated that science couldn’t be limited to what is known with certainty. Therefore, logical positivism (e.g., Carnap, 1934) took as its main task the establishment of criteria for what would constitute rational scientific inquiry, without regard for how scientists actually conduct their research (Garrison, 1996). In other words, the logical positivists sought a solution to the demarcation problem by defining and delimiting scientific logic.

When the approach to demarcation of the logical positivists was also found by scientists to be an inadequate guide, Karl Popper (1959) attempted to differentiate science from pseudoscience in yet another way. He argued that science progresses by submitting its hypotheses and theories to tests with the potential to falsify them, while the hypotheses and theories of pseudoscience can not be falsified. Unfortunately, the falsification criterion of demarcation had trouble explaining why scientific theories are seldom discarded when one or a few investigations produce anomalous outcomes for them.

How successful have the various attempts been to demarcate science from nonscience or pseudoscience? Not very. In fact, the consensus view in philosophy of science is that all such efforts have failed completely (Gjertsen, 1989; Laudan, 1981). The issue is not that there is no difference between science and other thoughtful or creative endeavors, but rather that no one has yet devised a set of criteria that reliably distinguishes scientific from nonscientific practices. Contrary to the position of logical positivists, scientists and philosophers of science have been unable to reach consensus on what constitutes scientific logic or the scientific method (Laudan, 1983). It seems that science is recognized more by its discoveries than by whether its methods correspond to any formal standards. Generally, it appears that scientists are those who contribute new knowledge to the sciences, even when they employ unusual or unorthodox methods to do so. In fact, the breadth of what is usually considered scientific across the natural sciences, and their relatives in engineering and the professions, makes it probable that any attempt to narrowly define science is doomed to the failure of rejection by practicing scientists themselves (Laudan, 1983).

It is true that there are a few philosophers of science who still maintain that science can and should be demarcated from nonscience. Even these few (e.g., Fuller, 1985; Gieryn, 1983), however, generally advocate using a kind of jury system. They argue that in such a system the practicing researchers in a field have the right to label those among their peers scientists as part of a social phenomenon, without using any objective criteria of methodological form that demarcates their work from nonscience.

At times, the demarcation of science from non-science has even been a political strategy. The philosopher and historian of science Imre Lakatos (1978) has pointed out that the Catholic Church in the 1600s engaged in demarcation to label findings of heliocentricity in astronomy as pseudoscience and then forced Galileo to recant. He also recalled that the Soviet Union in the mid-1900s used demarcation to label Mendelian genetics as pseudoscience and then tortured and executed its practitioners.

The National Reading Panel chose to engage itself in the messy and so far unsuccessful effort to solve the demarcation problem. The members boldly assert that they have differentiated the small amount of scientific, objective, and rigorous reading research from the great quantity of reading research that fails to merit one or more of these lofty labels. It has been more than 30 years since such a claim would not have appeared naive to anyone familiar with philosophy of science.

Moreover, the Panel’s criteria can be applied to its own work, raising several difficult questions. Did the Panel conform to its own standards? By its demarcation

criteria, is its own work scientific? Did the members of the Panel operate in a scientific, objective, and rigorous manner when they chose their procedures for conducting their review of reading research? Unfortunately, the answer seems obvious. Where are the scientific, objective, and rigorous studies that compare different ways of selecting and reviewing literature to improve practice? Is there experimental or quasi-experimental evidence demonstrating the superiority of the Panel's approach to determining which studies are a better guide to practice? No, members chose their demarcation criteria on logical rather than empirical grounds. Alas, the NRP's demarcation criteria do not pass its own standard: The Panel members' determination of what reading research is scientific is not scientific, as they themselves define it.

### **Verificationism**

Ignoring how practicing scientists conduct their research, positivists of various stripes (old, logical, and neo) have privileged one or another brand of verificationism. For example, verifiability-in-principle was the criterion that the logical positivists employed to demarcate science from non-science (Ayer, 1946; Carnap, 1934). To them, the meaning of any statement was the method of its verification. That is, any statement, however tentative, that could not be empirically verified was neither right nor wrong, but meaningless. Had scientists listened to the logical positivists—fortunately, most did not—they would have stopped searching for the truth of any hypothesis they did not then know how to verify. While the criterion of verifiability-in-principle was eventually abandoned by almost everyone, a broader and more nuanced neo-verificationism still has a few adherents among philosophers today, principally Michael Dummett (1976, 1991).

Verificationism is always concerned with the meaning of statements rather than the nature of reality. It interposes a theory of knowledge and a theory of language between scientists and the objects of their investigation. Positivists want their a priori views of science and of scientific logic and language to dictate what can be known.

The National Reading Panel clearly holds a verificationist philosophy of science. It states that “To sustain a claim of effectiveness [any instructional practice], the Panel felt it necessary that there be experimental or quasi-experimental studies of sufficient size or number, and scope... and that these studies be of moderate to high quality” (p. 1-7). Notice that the emphasis is not on effectiveness, but rather on claims of effectiveness. The true nature of reading or reading instruction is less important to the Panel than the need to “sustain [“verify”] a claim” (p. 1-7) about it.

The Panel's positivism is strongly held. Because statements about reading development and instruction apparently have scientific meaning only to the extent that they are empirically and experimentally verifiable, even a review of the experimental research is “subjective” (p. 5) unless the findings of those experiments can be combined in “a formal statistical meta-analysis” (p. 5). In other words, the Panel holds both a verificationism about reading research and a metaverificationism about reviewing reading research.

The Panel's verifiable-by-experiment criterion is applied quite consistently throughout its examination of reading research. The language of the Report betrays no tentativeness about the Panel's criterion. When the Panel appears tentative, a careful reading reveals that this tentativeness is certainly not about the criterion:

It should be made clear that these findings do not negate the positive influence that independent reading may have on reading fluency.... Rather, there are simply no sufficient data from well-designed studies capable of testing questions of causation to substantiate causal claims. (p. 13)

In other words, when its criterion for verification (data it considers sufficient from studies it considers well designed) is lacking, no claim can be verified.

### **A critique of the NRP's philosophy of science**

Most researchers, at least in the natural sciences, are scientific realists rather than positivists (Marsonet, 1995; Weinberg, 1992). Scientific realists are empiricists who build theoretical models, attempt to represent ever deeper layers of previously hidden reality, and seek full and satisfying explanations in order to achieve a clear and comprehensive understanding of cause-and-effect relationships (Cunningham & Fitzgerald, 1996). Scientific realists conduct experiments when experiments are called for, but they never confuse their methods with the reality their methods are used to discover.

Like all positivism, the Panel's work reveals a desire for certainty and a willingness to engage in reductionism to achieve it. All positivists have been antirealists (Cunningham & Fitzgerald, 1996), apparently because they are uncomfortable with the wide and never-closing gap between our knowledge and our questions (Searle, 1995). Their strategy has been to increase their comfort by reducing the questions one is permitted to ask, and reducing the ways one is permitted to answer them.

Practicing scientists of reading should be embarrassed by the simplistic, old-fashioned, and generally discredited verificationism of the National Reading Panel. In its assertions about the relationship between causal claims and the need for experimental evidence, the Panel has unwittingly allied itself with the research arm of the U.S. tobacco industry, the Tobacco Institute, which has long argued that the Surgeon General or anyone else has

no right to claim that smoking causes cancer because the relationship is merely correlational (Giere, 1997). The efforts of the NRP to formally demarcate science in reading from pseudoscience may actually be dangerous. While the members of the Panel I know personally are unquestionably well intentioned, one can be forgiven for being less certain about the Congress that requested the Report. I fear the philosophy of science that begins and permeates the NRP Report may have a chilling effect on the funding, publication, and influence of all reading research that fails to follow the positivist methodological standards it prescribes for our field.

### ***The NRP's doctrine of research design***

The NRP Report should also be seen as a declaration of a particular doctrine of research design. By largely limiting itself to the examination of experimental and quasi-experimental studies of reading, the NRP echoes the raging battle between experimentalists and correlationists in the social sciences of the 1950s and early 1960s. Its repeated view is that “correlations tell us nothing about the direction or sequence of a relationship” (p. 3-10). In fact, the Methodological Overview of the NRP Report reads almost as if there had been an open copy of Campbell and Stanley’s (1963) work in front of each of the Panel members as they developed their methodological standards.

In 1956, Lee Cronbach (1957) addressed an audience at the meeting of the American Psychological Association (APA). Unlike the NRP, Cronbach was willing in the title of his talk (“The Two Disciplines of Scientific Psychology”) to refer to some of both experimental and correlational research as scientific. In his presentation, Cronbach famously called for a crossbreeding of experimental psychological research methods with those correlational methods used to investigate individual differences in psychology. This new genre of research came to be known as the study or science of Aptitude by Treatment Interactions (ATIs).

Eighteen years later, Cronbach returned to APA (1975) to discuss the state of the then-thriving subdiscipline of ATI research. Surely his comments were not what his audience had expected. After praising what ATI research, especially in instruction, had contributed, he stated that such research was no longer sufficient because “Interactions are not confined to the first order; the dimensions of the situation and of the person enter into complex interactions” (Cronbach, 1975, p. 116). Stepping back to evaluate the previous 30 years of research in psychology, Cronbach said that, “Taking stock today, I think most of us judge theoretical progress to have been disappointing” (p. 116). In this evaluation of research, including ATI studies, he especially noted the limitations of the “two-group experiment” (p. 116).

With courageous candor, Cronbach related how he and his coauthor, Richard Snow, had “been thwarted by the inconsistent findings from roughly similar inquiries” (Cronbach, 1975, p. 119) in their attempts to generalize from results of ATI studies on instruction. From this experience, he came to realize that untested interactions, especially of a higher order, can always be envisioned for any study. Then, in comments anyone today should find eerily prophetic, Cronbach questioned the eagerness of some social scientists of the time “to establish rigorous generalizations about social policy by conducting experiments in the field” (p. 122).

Cronbach (1975) did not conclude his remarks by opposing scientific psychology or calling for an end to experimentation. On the contrary, he expected both to continue and prosper. What he did call for was the end to simplistic and reductionist reporting of scientific research. If he were to make the same talk today, surely he would castigate the reporting of nothing but effect sizes with the same fervor he expressed then against the reporting of “nothing save F ratios” (p. 124). What he endorsed instead was “the scientific observation of human behavior” (p. 124) with an emphasis on descriptions. In opposition to purely numerical products of research, he cited Meehl (1957) to agree with him that “we [scientists] have to use our heads” (p. 126).

Whether either of them were present to hear, or later read, Cronbach’s (1975) remarks, Jay Samuels and David Pearson worked to establish a similar spirit of broadened and balanced inquiry in our field during their editorship of *Reading Research Quarterly* from 1979—1985. Early on, they expressed an appreciation for the strengths and limitations of both experiments and naturalistic observation and called for the recognition of “the symbiotic relationship between paradigms” (Pearson & Samuels, 1980, p. 430). Later in his tenure as coeditor, Samuels (1984) echoed Cronbach’s concern with complex interactions that make it impossible to expect experimental science to find simple, all-embracing laws that generalize. He then discussed the implications for reading instruction of overlooking such interactions:

Many of our educational pundits appear to believe there are universal approaches to instruction and development of curricular materials which will work for all children under all conditions. They seem to ignore differences in intelligence and home background conditions. Depending on these variables as well as the degree of motivation and prior knowledge brought to the task of learning to read, it is highly likely that some approaches to instruction should be better for some children and different approaches

should work better for other children. (Samuels, 1984, p. 391)

In light of this historical background, the experimentalism of the NRP reminds me of Rip Van Winkle. It is almost as if the Panel fell mysteriously asleep 20 years ago and awoke just in time to do what the Congress and the NICHD convened them to do.

### **Equating reading education with interventions**

The NRP maintains that “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” and states its belief “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” (p. 27; p. 1-5).

This argument is based on a metaphor of reading instruction being like the curing of psychological and physical diseases. The Panel’s unquestioned assumption of this metaphor has the regrettable effect of reducing schooling in general, and reading education in particular, to a series of low- or noninteracting interventions. What if healthy human development is a better metaphor for schooling and the teaching of reading, pre-K through Grade 5 and beyond, than is the metaphor of treatments for specific mental or medical ailments? This metaphor would not negate the need for intervention research when particular treatments for specific reading disabilities or particular short-term learning outcomes are tested, but it would certainly broaden the research base for “the teaching of reading and in prevention... of reading disabilities” (p. 27; p. 1-5) beyond that considered scientific, objective, and rigorous by the Panel.

The NRP’s findings relative to the value of systematic phonics instruction and attempts to increase independent reading illustrate the limitations of experimentalism as a doctrine of research design and treating ailments as a metaphor for reading education. To see the inadequacy, consider two possible claims one could make about reading instruction:

1. Systematic phonics instruction in first grade is a cause of better reading ability by fifth grade and beyond.
2. Increased independent reading in the elementary grades is a cause of better reading ability by fifth grade and beyond.

It is difficult to see anything unreasonable about either of these claims or anything unscientific about wanting to evaluate them.

Because of its doctrinaire experimentalism, however, the Panel chose to evaluate all allegations about the effectiveness of systematic phonics instruction and attempts to increase independent reading in ways that cannot serve to shed much light on important claims like the two stated above. First, the Panel limited the duration of the effect of instruction to the length of time between the official onset of the intervention and the final data collection in each particular study. Such studies of the effects of smoking would be far less threatening to the tobacco companies than the devastating studies of longer term effects have been. Indeed, it may be the long-term and complex nature of reading development, and indeed of all schooling, that makes the NRP’s experimentalism most questionable.

Second, the Panel members forced themselves to attempt to select one or a few dependent variables that would permit them to conduct a meta-analysis or, at least, a “subjective qualitative analysis” (p. 5). So they tried to measure the short-term value of systematic phonics instruction using a reading comprehension dependent measure. Equally oddly, they tried to evaluate the short-term value of increased independent reading using a fluency dependent measure. Surely, these are examples of trying to pound square pegs into round holes because someone decided a priori that it would be easier to compare only round holes with one another.

What research designs would be more appropriate if healthy development were a better metaphor for learning to read than treating a range and sequence of diseases? They would be designs that test aspects of sophisticated theories of reading development. Wouldn’t it have made much more sense for the Panel to attempt to test one or more theories of reading development that endeavor to come to grips with the long-term and interactive nature of schooling? Why not, for example, identify a theory or model of reading or of reading development that includes a complex causal network? I have argued for example, that decoding by phonics has only small direct causal value for silent reading comprehension, but that it has important indirect causal value (Cunningham, 1993). That is, decoding by phonics contributes directly to the acquisition of automatic word recognition, which, in turn, has direct causal value for silent reading comprehension. This aspect of my model has much research to support it (e.g., Share, 1995), but it is difficult to imagine an experiment or quasi-experiment that would last long enough to conclusively test this indirect yet still causal relationship. Even if such an experimental study has been or could be done, it is a real stretch of the imagination to expect enough of them to make a meta-analysis possible. Yet do we want to ignore, or leave untested, theories that posit long-term, indirect causal relationships between decoding by phonics and ultimate reading comprehension

ability, or between world knowledge, wide independent reading, and ultimate attitudes toward reading, self, and school?

When the Panel equated reading education with a series of interventions, it made a fatal error our field can not afford to accept. It seems especially ironic that it made this error in the name of an organization given to the study of health and human development.

### **A critique of the NRP's doctrine of research design**

I contend that education, including the teaching of reading, is more like fostering healthy human development, building a successful business, maintaining an effective military, and providing good parenting than it is like administering medical or psychological interventions. American business and the American military are each the envy of the world, yet imagine how little of their cumulative wisdom and common practice is supported by the kind of research the NRP would insist upon for investigating claims about reading instruction. For instance, what would happen if parents began to feel doubts about any practice that does not have enough experimental support to conduct a meta-analysis?

Get intelligent people together as a committee and sometimes they collectively act with less common sense than any individual among them has. The experimentalism held so unwaveringly by the NRP violates all common wisdom. Such a doctrine will not do in reading education and must not go unchallenged.

### **The NRP's findings and determinations**

The members of the Panel divided themselves into five subgroups, with several members serving on more than one. These subgroups each examined the experimental and quasi-experimental research on the five main topics they had chosen: alphabeticity, fluency, comprehension, teacher education and reading instruction, and computer technology and reading instruction. In this section, I will briefly review the findings and determinations of the five subgroups. Before doing so, however, it is important to consider whether such a review is even necessary after critiquing the Panel's philosophy of science and doctrine of research design that guided all five subgroups in their work.

If the Panel's philosophy of science and doctrine of research design are seriously flawed, as I have argued, does that mean its findings are inevitably also flawed? Positivists and other antirealists would think so, because they hold that reality is always determined by the methods and language employed to examine and interpret it. Scientific realists do not concur. We agree that a misunderstanding of science or a limited approach to research design will inevitably lead to some mistaken or limited findings, but not all findings will necessarily be mistaken or limited. In the case of the NRP Report, it may be that some or even all the findings of the Panel happen to be what would have been found had members approached their job differently. Therefore, I conclude that the findings of the Panel still need to be evaluated on their likelihood to conform to reality given a broader view of epistemology, a more versatile set of research tools, and a different metaphor of reading education.

Other questions also require an analysis of the Panel's results and interpretations. What is the relationship between the Panel's approach and its findings? How consistently did Panel members apply their own standards when they conducted their selection, analysis, and interpretation of literature on reading instruction and development? These questions can be answered only by a review of the actual findings and determinations of the subgroups.

### **Alphabeticity**

The word *alphabeticity* is utilized by the Panel to group and label research on the topics of phonemic awareness (PA) and phonics instruction. The two topics are dealt with separately, with little explicit discussion of the relationship between them.

*Phonemic awareness.* The PA training that the Panel finds most effective is 5 to 18 hours of explicit and systematic, small-group instruction with one or two tasks of manipulating phonemes with letters, given to preschool and kindergarten children. Because the recommended instruction is "with letters" (p. 8; p. 2-4), the Panel's finding is tantamount to endorsing systematic phonics instruction in preschool and kindergarten (Yopp & Yopp, 2000). How should we, as a field, react to such a recommendation?

It does seem to me that, at the present time, the burden of proof (Giere, 1997) is on those who would have us do nothing instructional to foster the development of children's phonemic awareness. I believe we now have enough evidence that phonemic awareness is a necessary component of learning to identify words and that it is lacking in enough learners so we, as a field, must not leave its acquisition to chance. Had the Panel stopped there, I would endorse the finding wholeheartedly.

I also contend, however, that the burden of proof at this time is on those who would standardize PA training when so many questions about it remain unanswered. The chief question is the one that the Panel largely ignored throughout its entire work, even including its calls for future research: What are the long-term effects on silent reading comprehension ability, the reading habit, and attitudes toward reading, self, and school of its recommended changes in early reading instruction?