

B | BROWN CENTER LETTERS ON EDUCATION

Reading Second

By Grover "Russ" Whitehurst

Dear Colleague,

Former President George W. Bush finished his tenure without having won congressional renewal of his No Child Left Behind (NCLB) policy, which in a seven-year span has increased nationwide tracking of student achievement. With President Barack Obama now at the helm, NCLB is up for debate.



Above: Children raise their hands in a classroom.

A key component of NCLB is Reading First (RF). Created to encourage the use of scientifically based research as the basis for reading instruction, RF aims to ensure that all children learn to read well by the end of third grade. Reading is a foundational skill that affects the rest of a child's opportunities for learning from the earliest grades in school. However, the scientific evidence for RF was weaker than acknowledged at the outset. Now, after six years of implementation, the program has not met its lofty goals.

Where do we go from here? Program advocates want to stay the course on the assumption that more time will be sufficient for RF to realize its potential. This is a leap of faith. Opponents would roll funding for reading instruction into the Title I block grant for education of the disadvantaged, with minimal federal strings—a failed tactic of the 1990s.

There is another way: Extend the state-based standards and assessment provisions of NCLB down to grades K-2 for reading, while removing federal specifications for reading instruction. The strongest pillar of NCLB is its requirement for aligned standards and assessments coupled with public reporting of results. Yet, current requirements for reading start at grade 3, which is too late for the nation's weakest readers.

A requirement for clear state or regional standards for what children should accomplish in reading, starting in kindergarten, along with annual assessments aligned to those standards, should be a core reform of the next version of RF. Those standards should include not only the ability to translate print into speech but also the vocabulary and background knowledge that children need in order to comprehend what they read. More investment in R&D on reading instruction and better mechanisms for encouraging utilization of research findings in the classroom are also needed. Each is an appropriate federal role and a promising basis for the

design of the second version of Reading First.

You can download the full report below. I look forward to a robust debate on Capitol Hill and in the education community.

Regards,

Russ Whitehurst
Director, Brown Center on Education Policy.

Background

The No Child Left Behind Act of 2001 (NCLB) was a signature domestic initiative of the Bush administration. Reading First (RF) is a key component of NCLB, intended to ensure that all children learn to read well by the end of third grade. RF is remarkable in the degree to which it prescribes from the federal level the framework for reading instruction that is to be used in local school districts. In the rest of NCLB little is said about what in particular students should know, and outside of RF the bill is assiduously quiet on how students should be taught.

The architects of RF justified its intrusiveness on the basis that the prevalence of reading problems, particularly among children from low-income and minority backgrounds, was equivalent to national health emergency that existed because the findings from research were not being applied in the classroom. Just as the passage of national laws banning smoking in public places could be justified because of the strength of the evidence linking cigarette smoke to cancer, so too could a law dictating how reading is to be taught in public schools be warranted on the basis of scientific evidence on how children learn to read.

The Evidence on Reading First

There are two large empirical flies in the RF ointment. First, there was an absence of evidence going in that a large-scale program with the features of RF would eliminate reading difficulties by the end of third grade. Second, there is substantial evidence after 6 years of implementation that RF has not had much of an impact on children's reading comprehension.

On the first issue, the scientific evidence for the effectiveness of the type of instruction favored by RF was derived from a scientific literature of studies that in the parlance of health care research would be characterized as pre-clinical or early stage trials. In other words, these were studies that implemented reading interventions in small samples of schools under conditions that were tightly controlled by university-based researchers. Such studies are capable of demonstrating that an intervention can work under hothouse conditions, but not that it will work when scaled-up. Large, multisite field trials under ordinary conditions of use are required for that and were largely absent in the run-up to RF. Further, the science behind RF was on the firmest ground when it came to how to teach decoding (the ability to translate print into speech sounds) to struggling first graders. It was much weaker when it came to the other grades, to vocabulary, and to comprehension, which is the end game in reading. Finally, the research studies on which RF was built were focused on individual elements of what came to be RF, e.g., teaching phonemic awareness. RF cobbled the insights from this research into requirements for an omnibus intervention in five different domains of reading instruction, coupled with requirements for particular curricula materials, assessments, and teacher professional development. The recipe for this was untried prior to being written into law. Thus the science for the actual intervention prescribed by Congress was substantially weaker than acknowledged at the time.

The gap between what was known from the science of reading in 2001 and the design of an effective nationwide reading program was surely as large as the gap between rocket science and a successful moon shot when President Kennedy announced the goal of a manned moon landing in 1961. Yet it took the nation 8 years and an investment of approximately \$140 billion in today's dollars to put humans on the moon and

bring them back. In between came a lot of incremental engineering and more than a few failures. Should we be surprised, then, that the demonstrable results of RF after just 6 years and \$6 billion of investment are less than its proponents expected?

The two largest and most rigorous studies of RF examined reading test scores for children in schools in the RF program vs. results from similar children in non-RF schools. The more carefully controlled of the two studies assured that all the differences between RF and non-RF schools other than the presence of the RF program had been accounted for (see: <http://ies.ed.gov/ncee/pdf/20094038.pdf>). Thus it allows strong causal conclusions about the impact of RF. It found an effect of RF on phonics-related word reading skills in first grade but no effect on reading comprehension at any grade. It is notable that this study found an effect of RF where the science was strongest and where the implementation of RF was focused.

Two potential drawbacks of this study are that it used a sample of 18 sites that may not have been representative of RF schools nationally, and it compared RF with non-RF schools within the same school districts. On the latter point, critics have argued that districts in the study may have encouraged their non-RF schools to engage in the same practices as their RF schools, thus lessening the possibility of finding differences in student outcomes between the two types of schools.

The second, methodologically weaker study compared gains on state reading assessments in all RF schools in 41 states with gains in all non-RF Title I (low-income) schools in the same states (see <http://www.ed.gov/rschstat/eval/other/readingfirst-final/readingfirst-final.pdf>). Because the RF and non-RF schools in this study cannot be assumed to be equivalent in terms of student characteristics and other factors, as they could in the rigorous study, a finding of a substantial difference in reading gains favoring RF schools would have to be interpreted cautiously. However, substantial differences were not found. In fact, there were no significant differences in student reading gains between the RF and non-RF schools in a majority of states (23/41), and small differences favoring RF in the remainder. The study's authors describe the overall difference in student outcomes between RF and non-RF schools as "small in magnitude".

Here are two large studies with very different samples and analytic approaches. The strengths and weaknesses of each study are largely offsetting. Both converge on the same conclusion: RF has had, at best, small effects on student reading achievement.

What are the possible futures of Reading First?

Stay the course

Congress would continue to mandate particular forms of reading instruction in the classroom and the U.S. Department of Education would be responsible for implementing and enforcing those requirements in thousands of school districts across the nation.

The argument for sticking with what we're doing is that large-scale interventions in schools may take longer to take root than has been given RF. Proponents of this view point to Alabama, where a RF-like initiative was instituted in 1998. Alabama's reading scores on the National Assessment of Educational Progress did not start to rise until almost a decade later. Perhaps RF will begin to generate larger effects in another few years.

One problem with extrapolating the Alabama experience to RF is that the Alabama program had not expanded its reach into many schools that participate in the National Assessment until late in the decade in question. Thus, to the extent that the late rise in test scores is attributable to the Alabama reading initiative, we don't know whether that is because it took schools in the reading initiative many years to make their programs successful or whether the programs were successful early but didn't affect National Assessment scores because they weren't represented in the sample of schools that participated in the assessment.

The first of the studies described previously found declining differences in instructional practices between RF and non-RF schools over three years of implementation of RF. This isn't consistent with RF getting better with age. Further, other large scale school reform efforts have shown an impact on reading achievement after two or three years of implementation whereas RF has been in place for six years with only limited effects. Thus there is little evidence to support the view that all RF needs is patience.

Cut and run

Another possibility is for Congress to abandon any targeted appropriations or special requirements with respect to reading, thus returning to the status quo prior to passage of the Reading Excellence Act of 1998 (the predecessor to RF). The funding currently appropriated for RF would be rolled into Title I block grants, which school districts could spend on RF-like programs if they wished. The House of Representatives, in reaction to concerns about the administration of RF, moved in this direction in 2008 by zeroing out RF appropriations, which were later restored to two-thirds of their previous level.

The problem with cut and run is that reading is too important to be a discretionary investment of federal funds. Reading achievement among fourth graders has gone up a few points on the National Assessment since 1992. But in 2007, the last year for which results have been reported, 50% of fourth graders from low-income families scored Below Basic. Such children would typically not be able to demonstrate an understanding of the overall meaning of what they read or make relatively obvious connections between the text and their own experiences. Their educational lives are at risk as a result. There is simply no warrant in these abysmal results for going back to the good old days of reading instruction when student achievement was at even lower levels.

Extending standards and accountability to early reading

The strongest pillar of NCLB is the requirement for standards for what students should know at particular grades, annual assessment of students against those standards, and public reporting of results. NCLB was built on the standards and accountability movement that began in some states in the early 1990s and is a key feature of many education systems around the world.

It is difficult to conduct unimpeachable research on the effectiveness of standards and accountability systems because there is no possibility of comparing states or countries that differ only in the presence or absence of that component of an education system. Nevertheless, a number of studies that have applied the strongest statistical approaches to the available evidence converge on the conclusion that academic achievement is higher when there are external examinations that are guided by expectations of what students should know and be able to do.

Perhaps we can get further in enhancing reading achievement at the federal level by relying on standards and accountability to drive reform than by dictating the details of reading instruction in legislation.

Standards and accountability can have a direct impact on learning if they motivate schools or students to devote more time to a topic than would otherwise be the case. This is captured in the aphorism, what gets tested gets taught. This effect is probably largely responsible for dramatic increases in student mathematics achievement on the National Assessment going back many years. For example, average performance on the National Assessment in math at fourth grade rose 14 points between 2000 and 2007, which is a difference that represents well over a year's worth of schooling. There were no particular innovations in mathematics instruction during that period, and no Math First, so these improvements are likely due to more student and teacher time on math tasks that have been identified as important in a set of state and national assessments and standards. These standards and assessments have become quite similar to each other and well known by math educators.

Why have standards and accountability alone seemed to work for math, while standards and accountability plus RF have produced only small effects for reading? There are two reasons.

First, it is likely that grade three, which is the point at which NCLB requires state standards and annual testing for reading and math, is fine for math but too late for reading. Reading is a foundational skill that affects the rest of a child's opportunities for learning from the earliest grades in school. If a child has reading difficulties that aren't remediated in the beginning grades the child will fall further and further behind.

Mathematics is important too. But it is far more compartmentalized in the first years of elementary school than reading and it is easier for schools to make up for early deficiencies. Thus, weak performance by a school's students on a third grade assessment of mathematics may motivate that school to beef up the intensity of its math instruction, and that may be enough to raise performance in subsequent years. The same result from a third grade assessment of reading may motivate as much, but the wherewithal for the school to remediate reading problems at that point in a child's schooling will be lacking.

RF, which covers grades K through three, has generated a de facto standards and accountability system for those grades, but it is seriously incomplete. The one assessment that ties together most RF schools is DIBELS, a measure of children's ability to decode printed letters and read connected text fluently. Student performance on oral reading fluency is measured by having students read a passage aloud for one minute. The number of correct words per minute from the passage is the oral reading fluency score. Thus DIBELS measures the ability of a child to automatically translate print into speech. It does not measure the child's understanding. Fluent decoding of print is necessary but not sufficient for reading comprehension. There are many children who decode fluently but who struggle to comprehend.

RF programs nationally are bound together by DIBELS. The assessment provides clear standards for decoding fluency in the form of norms for each grade, the professional development of teachers within RF has a focus on how to teach decoding, and DIBELS scores are typically reported for the purposes of mandated state reports under RF. Further DIBELS is typically administered several times a year so that the progress or lack thereof of individual students can be attended to by teachers. Recall that the more rigorous of the two national evaluations of RF found a significant impact of RF on decoding skills, so this approach has worked.

However, when it comes to the other areas that are critical to reading comprehension such as vocabulary and world knowledge, there is no such clarity. Most states require a year-end test of comprehension in RF, but that test is most likely to be a standardized assessment such as the Iowa Test of Basic Skills or the Terra Nova rather than an assessment that is aligned with specific learning standards. Further, results from these nationally normed, generic assessments of reading comprehension, given at the end of the school year, do not generate feedback to teachers that can help them identify which students are struggling and need more help. When it comes to knowing what content vocabulary or world knowledge a second grader should have mastered by the end of the year, most states and the federal government leave schools and teachers in the dark. If teachers don't know what they're expected to teach, they won't teach it. If schools and districts don't have public accountability for what they are teaching they will give a short shrift.

The second reason that standards and accountability have worked for math but not for reading is that reading depends much more on background knowledge acquired in the home than does math. And, most critically, the research base on how schools are to make up for differences in family background that affect reading is relatively weak.

We know from the existing evaluations that teachers in RF classrooms spend a considerable portion of the reading block teaching comprehension, but the approach favored in RF is to teach children comprehension strategies, e.g., summarizing the key points in text. Comprehension strategies can be useful, but children also need knowledge of the world to read for meaning. Imagine reading a newspaper headline: "Rupert Murdock Buys the New York Times." Using the comprehension strategies taught in RF might make you more likely to remember who bought the Times or what Murdock bought, but without broader knowledge of the Times, Murdock, and the news business, you wouldn't understand what was newsworthy. Many activities carried out by middle class parents such as conversation at the dinner table, visits to museums, and exposure

of their children to informational forms of media generate core knowledge of the world that enhances reading comprehension.

Well formulated reading standards and aligned assessments for grades K-2 that incorporate the vocabulary and world knowledge that children need to comprehend the material they will be expected to read would increase instructional time in those areas. That alone is likely to have an impact.

At the same time we need to acknowledge what was swept under the rug in RF: that we don't have a lot of evidence on how teachers should build world knowledge, vocabulary, and related skills such as proficiency in analysis of text. Good standards and assessments in these areas, absent in RF, would motivate the research and development enterprise to create a new generation of programs and practices that are effective in imparting this knowledge to children in a school setting. It will take time to engineer more coherent and effective approaches in these areas. We will need to learn, engineer, and evaluate as we go.

The federal government has a critical role to play in sponsoring research and development on effective ways to teach reading and in providing incentives and mechanisms for translating that research into practice. If it plays that role effectively in education, as it has in health care, and pushes states to develop strong standards and assessments for what children should know and be able to do in reading from the earliest grades, we will see progress.

Recommendations to Congress and the administration for Reading Second

- **Continue a funding stream for beginning reading for which states will have to compete.**
Reading is too important to allow states and schools to decide whether it should be a priority for the use of federal funds.
- **Raise the funding level.** \$1 billion a year is insufficient to the task.
- **Remove requirements for instruction in the five essential components of reading and other specifications that try to translate research findings into instructional mandates.** Congress doesn't tell health care professionals how to practice and it shouldn't tell educators how to teach. Even if the research were definitive, which it isn't; and Congress could translate it into legislation without distortion, which it can't; and bureaucrats in the U.S. Department of Education could implement it unimpeachably, which is unlikely, science is dynamic. We shouldn't accept a process that requires Congress to rewrite legislation in order to bring teaching practice in line with evolving research findings.
- **Require states that compete for reading funding to establish learning standards for reading in grades K-2, administer annual assessments aligned with those standards, and report disaggregated results at the school level.** We need good standards and accountability for the grades in which reading instruction is critical in order to drive instructional priorities and R&D.
 - **Require states to measure and report school-level results of the state beginning reading assessments in terms of growth as well as end-of-year status.**
 - **Do not impose proficiency targets and sanctions on schools. Let the public reporting of results and public recognition of exemplary performance drive accountability.**
 - **Provide incentives to states either to work as consortia to develop standards and assessments or to adopt existing high quality instruments** (see below). This will be more efficient and is likely to result in better and standards and assessments than having each state go it alone.
 - **Authorize and provide funding for the Institute of Education Sciences within the U.S.**

Department of Education to fund competitively no less than two entities to provide model standards for beginning reading and assessments aligned to those standards.

These standards and assessments would be available for any state that wished to adopt or adapt them. They would not be national standards and assessments, which is one reason to fund at least two sets, and no one would be required to adopt them.

- **Provide a new and dedicated funding stream to support research and development on practical school-based strategies for enhancing the vocabulary, background knowledge, and text analysis skills that children need for reading comprehension, and to disseminate the results of that R&D through easily accessible and effective materials for practitioners.** We know less than we need to know about how teachers and schools can make-up for the large differences among children in vocabulary and world knowledge that are associated with family background.
- **Expect more than has been delivered by RF but less than uniform reading proficiency by the end of third grade.** Reading cannibalizes everything that children know. Children from families in which adults have low levels of education have disadvantages that will be reflected in reading comprehension. With good standards and accountability for beginning reading, more investment in R&D on effective reading programs and practices, and mechanisms for translating research findings into practice, we can do substantially better than we are doing now.

Author

Grover “Russ” Whitehurst is a senior fellow and director of the Brown Center on Education Policy at the Brookings Institution. Previously, he was the director of the Institute of Education Sciences in the U.S. Department of Education.

About the Brown Center on Education Policy

Established in 1992, the Brown Center on Education Policy conducts research and provides policy recommendations on topics in American education. The Brown Center is part of The Brookings Institution, a private nonprofit organization devoted to independent research and innovative policy solutions. For more than 90 years, Brookings has analyzed current and emerging issues and produced new ideas that matter - for the nation and the world.

About the Brown Center Letters on Education

The Brown Center Letters on Education will be published several times each year. Each Letter will address an issue that is of current interest to education policy makers or practitioners by providing background, reporting relevant evidence, and making recommendations for action. Authors of Letters will be scholars with experience bridging the worlds of research and practice. Their charge is to make specific and practical recommendations that are grounded in a careful and dispassionate consideration of evidence. While the recommendations are to be grounded in evidence, they will rarely be compelled by it. Other scholars looking at the same evidence could generate different interpretations.

Brown Center on Education Policy

The Brookings Institution
1775 Massachusetts Ave. NW
Washington DC, 20036
202.797.6090
202.797.6144 (f)
www.brookings.edu/brown.aspx

This publication was made possible in part by a grant from the W.T. Grant Foundation. The statements made and views expressed are solely the responsibility of the author(s).