

Did Reading First Work?

By Stephen Krashen, University of Southern California

October 2006

The recent report from the Office of the Inspector General (2006) concerning wrongdoing in awarding Reading First grants has resulted in a defense of Reading First itself. While condemning the violations in administering Reading First, Secretary of Education Margaret Spellings (2006) insists that Reading First has been successful, citing three pieces of evidence: A rise in NAEP scores, a study from Michigan State, and test scores from the State of Washington. The International Reading Association has also strongly denounced the unethical practices described in the report, but notes that two studies provide evidence of the program's effectiveness, "Keeping Watch on Reading First," from the Center for Educational Policy, and the Reading First Implementation Evaluation.

None of these five sources provides any convincing support for Reading First.

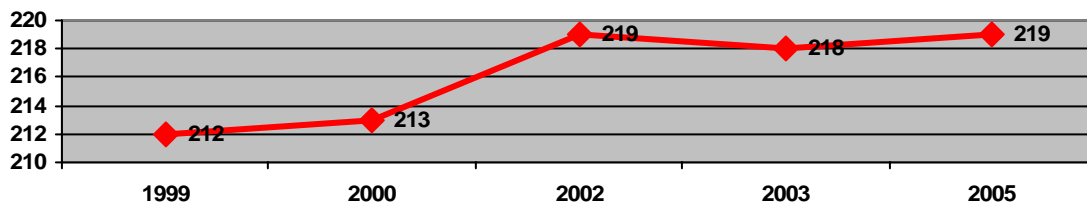
NAEP Scores

According to Spellings (2006), "The long-term trend data from the National Assessment of Educational Progress (NAEP), indicate that over the last five years, more reading progress has been made among nine-year-olds than in the previous 28 years combined ... I believe this is due in part to the contributions of Reading First and other programs under the No Child Left Behind Act."

The crucial question is whether there has been an increase since NCLB and Reading First have been implemented.¹ The five year trend analysis Spellings mentions (National Center for Educational Statistics, 2004) provides scores only for 1999 and 2004. There are no trend scores for the years in-between. There was indeed an increase, from 212 for fourth graders in 1999 to 219 in 2004, but as Bracey (2006) has pointed out, NCLB and Reading First were not introduced until 2002-2003. From these scores, we cannot determine whether these programs deserve the credit.

Regular ("main NAEP") scores, in fact, suggest that the jump happened before NCLB and Reading First, between 2000 and 2002 (NAEP, 2005).

NAEP Grade 4 Reading Scores, 1999 - 2005



¹ For discussion of NAEP scores before 1999 and Spellings' claims, see Bracey (2006).

Regular NAEP scores are not considered appropriate for comparisons; only trend scores are. But this data suggests that NCLB and Reading First were not responsible for gains between 1999 and 2005. (Note that the trend and regular scores were the same in 1999, and the 2004 trend score is very close to the 2003 score and identical to the 2005 score.)

Defenders of NCLB also claim that the “increase” is “especially true for groups that have lagged far behind in the past” (Shanahan and Hynd-Shanahan, 2006). Again, examination of NAEP scores does not support this. The differences in NAEP reading scores between children eligible for free or reduced lunch and those not eligible are nearly the same in 2005 as they were in 2003 (NAEP, 2005).

For fourth grade reading:

2003: high poverty mean = 201, low poverty = 229; difference = 28

2005: high poverty mean = 203; low poverty = 230; difference = 27.

Again, these are regular and not trend scores and must be considered suggestive, but there is no evidence that NCLB has been of special benefit to low-income groups.

Bracey (2006) also notes that it is very unlikely that many Reading First children were included in the NAEP assessments in 2004 (and even 2005). NAEP is given to nine year olds, but Reading First is directed at grade three and lower. Many Reading First programs did not begin until late in 2003; in fact, Bracey notes that the application package for Reading First was not available until April, 2002.

In addition, according to the Center for Education Policy (2006), only 6% of public schools participate in Reading First. NAEP scores are considered to representative of all children in the U.S.

The Michigan State Study

Spellings (2006) notes that “A study by researchers at the University of Michigan showed Reading First students in that state are making continuous gains across the board from year to year.”

The study she is referring to is Carlisle, Cortina, Zeng, and Schilling (2006). Carlisle et. al. reported gains for children in 108 schools in Michigan during the second year of Reading First. That is, more children read at “grade level” (40th percentile) and fewer at an underachieving level (20th percentile) after two years of Reading First than after one year. The study was limited to grades one to three. Thus, the gains claimed are not “from year to year” but from only one year to the next.

Here are some typical results. For second graders, after one year in Reading First, 46% scored at the 40th percentile on the IBTS reading comprehension test. One year later, this increased to 53%., a seven percent gain. Similar gains were found for vocabulary and for word analysis subtests.

There are very serious problems with this study.

No control group. Carlisle et. al. say that it was impossible to find a control group because none were available matching the Reading First schools on levels of reading and levels of poverty (p. 5). The fact that there was a practical constraint, however, does not bestow validity on the study.

Test score inflation. Gains are only reported for the second year of Reading First. Gains after one year are not reported because, according to Carlisle et. al., “no baseline data prior to RF are available” (p. 7). This suggests that prior to Reading First, the ITBS was not given in Michigan at these grade levels. This appears to be the case (see e.g. Mackinac Center, 2002) Thus, we are dealing with a new test, at least new to these teachers and students, which means one can expect the improvement one typically sees after new tests are introduced, a result of increased teacher and student familiarity with the test. Typical test score inflation is about 1.5 to two percentile ranks per year (Linn, Graue, and Sanders, 1990; table 2, p.12), and increases in the percentage of children achieving at national norms typically rises one to five percent, similar to that seen in the Michigan data (Linn et al., p. 11).

No raw scores. Carlisle et al. do not report raw scores, but only provide the percentage of children who reached the 40th and 20th percentiles. Knowing that reaching certain percentiles is the target, it is common knowledge that schools tend to focus on students scoring just below these levels. Also, there is also no reason NOT to report raw scores.

Sample size not provided. Carlisle et .al. do not report precise sample sizes. We do not know if more or fewer children were tested on the posttest. There is no reason not to report this data and it is important to do so to avoid the suspicion that selective testing was taking place to artificially inflate scores.

Misleading effect size calculations. Carlisle et al. report large effect sizes for the gains, giving the impression that increases were truly huge. Their effect size calculations were based on the ANOVA results. A more obvious way is to use pre and post-test means and standard deviations, which would result in effect sizes about one-third as large as those reported. Even so, these would be based on percentages of students reaching certain levels. Use of raw scores and standard deviations are much more appropriate for effect size calculations.

No detail on implementation. It is difficult to describe practices in 108 different schools. Nevertheless, no detail is provided as to the nature of the Reading First intervention.

This study comes nowhere close to meeting the Department of Education’s own standards for scientific research.

Washington State

The third claim by Secretary Spellings is that Reading First students in the state of Washington showed a 22% gain “after the program had been implemented for two years.”

Inspection of the Reading First section of the State of Washington Department of Education website revealed only one report on test scores, a bar chart, showing the percentage of children that “met standard” and that “exceeded standard” in 1997 (all students), and the results for a 51 school group that were in Reading First in 2003 and 2005 (<http://www.k12.wa.us/curriculuminstruct/reading/readingfirst/pubdocs/ReadingFirstEthnicBreakdownDOE.pdf>).

In 1997, 24.5% met or exceeded the standard. For Reading First children in 2003, 39.7% met or exceeded the standard, and in 2005 the figure increased to 62.3%, a spectacular gain of 22.6%, a far greater yearly gain than that seen between 1997 and 2003.

Once again, however, there are flaws.

No comparison group. Scores for children not in Reading First in 2003 and 2005 are not provided.

No details about the test, number of children tested, SES, grades. We are not told what test was used, whether it was a new or old test or how many children took it each time. Breakdowns are provided by ethnic group, but not by the all-important levels of SES. We are not even told what grades were involved.

No details on implementation. As was the case with the Michigan study, zero information is provided on implementation.

It is remarkable that the Department of Education, so committed to “scientific” research and rigorous methodological standards, would even consider these sources of data. In fact, a glance at the standards used by the Department of Education to include studies in the “What Works Clearinghouse” (www.whatworks.ed.gov) shows that none of these “studies” would even be considered for inclusion.

Keeping Watch on Reading First

In a post on the International Reading Association web site, IRA executive director Alan Farstrup notes that “the Independent Center for Education Policy report, ‘Keeping Watch on Reading First,’ concluded that many states have found that Reading First funding has been critical to their progress.”

The most important aspect of “Keeping Watch” is that it provides no actual data on the effectiveness of Reading First. It consists only of the “... views of state and district officials” (p. 5). In addition, the study contains some curious omissions.

Responses from state officials

“Keeping Watch” reported that officials that in 19 out of 35 states “that reported reading was improving,” said Reading First was an important cause of the increase. Thus, a little more than half (54%) gave credit to Reading First for improvement. But Reading First has been implemented in all 50 states. Apparently, reading was not improving in the 15 other states. The data could be interpreted to mean that Reading First was thought to be helpful in only 15 out of 50 states – less than one-third

Responses from district officials

As was the case with the states, “Keeping Watch” only mentions districts that reported gains. Of these, 97% credited Reading First: “... the majority of districts view Reading First as effective and as broad reaching” (p. 8). Once again, how many districts did not report gains? The report claims to have surveyed 1,717 districts, but provides no breakdown was provided into those who gained and those who did not. And the district reactions are far more enthusiastic than reactions from state officials.

The best we can say about this report is that the results are mildly suggestive. One also wonders why “Keeping Watch” didn’t ask the real experts, the teachers.

The Reading First Implementation Evaluation

Farstrup notes that this report (Moss, Jacob, Bouley, Horts, and Poulos, 2006) “indicates that the program is having a positive impact with many states and localities.”

This report does not deal with student achievement at all, but focuses only on implementation and the relationship between Reading First and non-Reading First Title I reading instruction (Exec. Summary, p. 2). It is thus inappropriate to cite it as evidence for the success of Reading First.

What is of interest is the finding that teachers in Reading First schools reported spending more time on reading than those in non-Reading First Title I schools – 19 minutes per day more, or 100 minutes per week (p. 4). They also reported more use of supplementary materials (69% to 58%), among other changes in materials (p. 5). Since Reading First devotes more time to reading, if it is just as effective as what the comparison group does, it should appear to be better. In other words, Reading First should be better than doing nothing. But research so far does not even show this.

Conclusions

Reading First is based on the report of the National Reading Panel (NICHD, 2000). This report has been heavily criticized by a number of scholars who point out that there is insufficient evidence to support the National Reading Panel's claims that phonemic awareness training significantly improves children's reading, that the published research does not support the claim that systematic phonics instruction is superior to less intensive instruction, and that there is no evidence that skills-based approaches are superior to whole language. Also, contrary to the conclusions of the National Reading Panel, there is abundant evidence that encouraging children to read more in school is beneficial (see, e.g., Coles, 2003; Garan, 2002; Krashen, 2003; Allington, 2002).

A test of whether the Panel or the critics are correct is whether Reading First does or does not significantly improve reading. Thus far, Reading First has not been put to the test. Claims of its success or failure are premature.

References

Allington, R. (Ed.) 2002. *Big Brother in the National Reading Curriculum: How Ideology Trumped Evidence*. Portsmouth, NH: Heinemann

Bracey, G. 2006. Letter to Representative George Miller and Senator Edward Kennedy, September 25, 2006.

Carlisle, J.F., Cortina, K., Zeng, J., & Schilling, S.G. (2006). *Gains in Reading Achievement Over Two Years in Michigan's Reading First Schools*. Technical Report #3.1, Evaluation of Reading First in Michigan. Ann Arbor: University of Michigan.
<http://www.mireadingfirst.org/resources/research/downloads/tr0301.pdf#search=%22carlisle%20cortina%20zeng%22>

Center on Education Policy, 2006. *Keeping Watch on Reading First*. Washington, DC: Author.

Coles, G. 2003. *Reading the Naked Truth: Literacy, Legislation, and Lies*. Portsmouth, NH: Heinemann.

Garan, E. 2002. *Resisting Reading Mandates*. Portsmouth, NH: Heinemann.
 International Reading Association, 2006. IRA responds to Reading First report.
<http://blog.reading.org>

Krashen, S. 2003. False claims about phonemic awareness, phonics, skills vs. whole language, and recreational reading. <http://www.nochildleft.com/2003/may03reading.html>

Linn, R., Graue, E., & Sanders, N. 1990. Comparing state and district test results to national norms: The validity of claims that "everyone is above average." *Educational Measurement: Issues and Practice*, 10, 5-14.

Mackinac Center. 2002. Which education achievement test is best for Michigan? Mackinac Center for Public Policy. Policy Brief.
<http://www.mackinac.org/article.aspx?ID=4382>.

Moss, M., Jacob, R., Boulay, B., Horts, M. and Poulos, J. 2006. Reading First Implementation Evaluation: Interim Report. Cambridge, MA: Abt Associates

National Assessment of Educational Progress (NAEP), 2005. *The Nation's Report Card: Reading 2005*. Washington: National Center for Educational Statistics.

National Center for Educational Statistics, 2004. National trends in reading by average scale scores. <http://nces.ed.gov/nationsreportcard/ltr/results2004/nat-reading-scalescore.asp>

National Institute of Child Health and Human Development (NICHD). (2000). *Report of the National Reading Panel. 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*. [NIH Publication No. 00-4754]. Washington, DC: U.S. Government Printing Office. <http://www.nichd.nih.gov/publications/nrp/report.htm>

Office of the Inspector General. 2006. *The Reading First Program's Grant Application Process: Final Inspection Report*. ED-OIG/I13-F0017. Washington, DC: U.S. Department of Education.

Shanahan, T. and Hynd-Shanahan, C. 2006. A good start is not enough: What it will take to improve adolescent literacy. [TC Record.org](http://TCRecord.org). September 5.

Spellings, M. 2006. Memorandum to Jack Higgins, August 29. In Office of the Inspector General, 2006. *The Reading First Program's Grant Application Process: Final Inspection Report*. ED-OIG/I13-F0017. Washington, DC: U.S. Department of Education, p. 38.